

XX WPI; 2000-013229/01.
XX New peptide derived from murine epidermal growth factor (mEGF) -
XX Claim 4; Page 28; 35pp; English.
XX This is a peptide derived from mouse epidermal growth factor (EGF)
XX residues 33-42. This peptide is used in the invention to prepare a
XX composition to target laminin receptors. EGF derived peptides inhibit
XX blood vessel formation through their antagonism of the high affinity 67kD
XX laminin receptor found on endothelial cells. The peptide is modified
XX from the natural sequence to prevent protease attack. The peptide is used
XX in the preparation of a medicament for binding to laminin receptors as an
XX (ant)agonist. The medicament is also useful for healing endothelial cell
XX wounds and treating angiogenic diseases, especially retinopathy of
XX immaturity. Other diseases treated include metastatic cancer,
XX Candida spp. infection, and parasitic infestations like leishmania and
XX trichomonas vaginalis. The peptide are anti-angiogenic in human models.
XX The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,
XX and prevent tumour cell attachment to basement membranes.
XX Sequence 10 AA;

Query Match 100.0%; Score 54; DB 21; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.074; 0; Indels 0; Gaps 0;
Matches 9; Conservative 1; Mismatches 0;
QY 1 CVIGXSGDRC 10
DQ 1 CVIGYSGDRC 10
||||:|||||

RESULT 2
AAG66047
ID. AAG66047 standard; peptide; 47 AA.
AC AAG66047;
DT 27-FEB-2002 (first entry)
DE Mouse EGF motif sequence.
KW ErbB-4; neuregulin-4; NRG-4; pro-NRG-4; neuroprotective; vulnary;
KW cerebroprotective; vasotropic; antiparkinsonian; anticonvulsant;
KW cytostatic; nootropic; EGF; NRG-2.
OS Mus musculus.
PN WO200181540-A2.
PD 01-NOV-2001.
PF 20-APR-2001; 2001WO-IL00371.
PR 21-APR-2000; 2000US-0553769.
PA (YEDA) YEDA RES & DEV CO LTD.
PI Harari D, Yarden Y;
PI WPI; 2002-041398/05.
XX Novel ErbB-4 ligand, referred as neuregulin (NRG)-4 and polynucleotide
XX sequences encoding NRG-4, useful for upregulating or downregulating
XX ErbB-4 receptor activity to treat Alzheimer's disease, stroke, gastric
XX cancer
XX Disclosure; Fig 1c; 153pp; English.
XX The invention relates to a novel ErbB-4 ligand, neuregulin-4 (NRG-4).
XX NRG-4 binds to mammalian ErbB-4 receptor and can be expressed by standard
XX recombinant methodology. Pharmaceutical compositions comprising NRG-4 are

CC useful for regulating an endogenous protein affecting ErbB-4 receptor
CC activity in vivo. They are also useful for treating or preventing a
CC disease condition or syndrome associated with dysregulation of an
CC endogenous protein affecting ErbB-4 receptor activity, e.g., amyotrophic
CC lateral sclerosis (Lou Gehrig's disease), Bell's palsy, spinal muscular
CC atrophy, brain trauma, stroke, ischemia, Alzheimer's disease, Parkinson's
CC disease, epilepsy, multiple sclerosis, Huntington's chorea, Down's
CC syndrome, nerve deafness, neuropathy, muscular dystrophy, extramammary
CC Paget's disease, gastric, pancreatic, prostate, breast and ovarian
CC cancer, cervical carcinoma, endometrial adenocarcinoma, pancreatic D
CC cells-somatostatinoma and Zollinger-Ellison syndrome. The agent comprised
CC in the pharmaceutical composition includes a polypeptide (e.g., a soluble
CC ligand binding domain of ErbB-4 i.e., 1984; or a monoclonal, polyclonal,
CC humanized, single chain antibody or an immunoreactive derivative of an
CC antibody) capable of binding the endogenous protein affecting ErbB-4
CC receptor activity. Traceable synthetic/recombinant NRG-4-tagged molecules
CC can serve as a diagnostic tool in which cells binding NRG-4 can be
CC measured. Sequences AAG66044-53 represent the EGF-like motifs of various
CC growth factors.
XX

SQ Sequence 47 AA;

Query Match 100.0%; Score 54; DB 23; Length 47;
Best Local Similarity 90.0%; Pred. No. 0.29;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
DQ 33 CVIGYSGDRC 42
||||:|||||

RESULT 3
AAP40315
ID. AAP40315 standard; protein; 53 AA.
XX AAP40315;
XX 22-JUL-1992 (first entry)
DE Epidermal growth factor tripentacontapeptide.
KW EGF.
OS Synthetic.
FH Key Location/Qualifiers
FT Disulfide-bond 6...20
FT Disulfide-bond 14...31
FT Disulfide-bond 33...42
XX JF59027858-A.
XX 14-FEB-1984.
XX 05-AUG-1982; 82JP-0137128.
XX 05-AUG-1982; 82JP-0137128.
XX (NNSH) NIPPON SHINYAKU KK.
XX WPI; 1984-072465/12.
XX Synthesis of epidermal growth factor polypeptide - by condensation
XX of protected smaller peptide sequences, deprotection then oxidn. to
XX cyclise.
XX Claim1; Page 1; 8pp; Japanese.
XX The amino acid sequence is that of an epidermal growth factor
XX tripentacontapeptide which is synthesised by condensation of
XX protected smaller peptide sequences. This method produces the
XX peptide smoothly, with high purity and yield.

SQ Sequence 53 AA;
 Query Match 100.0%; Score 54; DB 5; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
 ||||:||||
 DB 33 CVIGYSGDRC 42

RESULT 4
 AAP91658
 ID AAP91658 standard; protein; 53 AA.
 XX
 AC AAP91658;
 XX
 DT 29-JUN-1990 (first entry)
 XX
 DE Amino acid sequence for naturally occurring epidermal growth factor
 (EGF).
 DE
 XX
 KW Epidermal growth factor; angiogenesis; synthetic peptide.
 XX
 PN W08901489-A.
 XX
 PD 23-FEB-1989.
 XX
 PF 10-AUG-1988; 88WO-AU00300.
 XX
 PR 10-AUG-1987; 87AU-0003629.
 XX
 PA (CSIR) COMMONWEALTH SCIENT ORG.
 XX
 PI McAuslan BR;
 XX
 DR WPI; 1989-068852/09.
 XX
 PT Synthetic peptide active in stimulating angiogenesis -
 PT has sequences corresponding to amino acid sequences occurring in
 PT epidermal growth factor.
 XX
 PS Fig 1; 1/1; 11pp; English.
 XX
 CC The inventors claim synthetic peptides which correspond to sequences
 CC occurring in EGF, but excluding EGF. The synthetic peptides
 CC correspond to active sequences from EGF fragments: 3-14, 12-14, 12-15,
 CC 3-10, 29-37 and 33-37 (Claims 2-7 and AAP91659-P91664 resp.). The
 CC peptides are angiogenic. Their relative shortness means that they pose
 CC fewer synthesis problems than the entire EGF molecule.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 54; DB 10; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
 ||||:||||
 DB 33 CVIGYSGDRC 42

RESULT 5
 AAR08007
 ID AAR08007 standard; protein; 53 AA.
 XX
 AC AAR08007;
 XX
 DT 25-FEB-1991 (first entry)
 XX
 DE Modified murine epidermal growth factor.
 XX
 PR Modified murine epidermal growth factor; stability; storage;

KW epithelial wounds; gastric acid secretion.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 11..11
 FT /label= E, N, Q, A, K
 XX
 PN W09013570-A.
 XX
 PD 15-NOV-1990.
 XX
 PF 09-MAY-1990; 90WO-US02600.
 XX
 PR 12-MAY-1989; 89US-0351773.
 XX
 PA (CHIR-) CHIRON CORP.
 XX
 PI Nascimento CG, Medina-Selby A;
 XX
 DR WPI; 1990-361427/48.
 XX
 PT Human epidermal growth factor - is substituted at position 11 for
 PT greater stability and improved storage life.
 XX
 PS Claim 9; Page 25; 32pp; English.
 XX
 CC The human rEGF is used to treat oversecretion of gastric acid or an
 CC epithelial wound. EGF is modified to increase its chemical
 CC stability. Its storage life is improved without diminishing its
 CC biological activity. The proteins may be prepared by traditional
 CC chemical or recombinant means.
 CC See also AAR08004.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 54; DB 11; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
 ||||:||||
 DB 33 CVIGYSGDRC 42

RESULT 6
 AAR67275
 ID AAR67275 standard; peptide; 53 AA.
 XX
 AC AAR67275;
 XX
 DT 01-AUG-1995 (first entry)
 XX
 DE Synthetic human/mouse EGF.
 XX
 KW Cell growth factor; viscoelastic solution; fibroblastic growth factor;
 KW EGF; epidermal growth factor; EGF; buffered solution; lubrication;
 KW carboxypropyl methylcellulose; cellulose gum; dextran; dextran sulphate;
 KW chondroitin sulphate; sodium hyaluronate; osmolality; mitogenic;
 KW wound healing; cell protection; cell coating; surgery; tissue space;
 KW hydroxypropyl methylcellulose; manipulation.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 37
 FT /note= "Given in the specification as Try"
 XX
 PN US5366964-A.
 XX
 PD 22-NOV-1994.
 XX
 PF 15-DEC-1988; 88US-0284533.
 XX
 PR 15-DEC-1988; 88US-0284533.

PR 13-NOV-1989; 89US-0434305.

XX (LIND/) LINDSTROM R L.
PA (SKEL/) SKELNIK D.

XX Lindstrom RL, Skelnik D;

XX WPI; 1995-005865/01.

XX Viscoelastic soln. contg. hydroxypropyl methylcellulose, sodium
PT hyaluronate, chondroitin sulphate and growth factors - used as a
PT surgical soln. to promote wound healing, e.g. of corneal cells.

XX Disclosure; Column 5; 8pp; English.

XX The sequences given in AAR67273-76 are cell growth factors which may
CC be used in the viscoelastic solution of the invention. The peptides
CC are derived from fibroblastic growth factor (FGF) and epidermal
CC growth factor (EGF). The viscoelastic solution also comprises a
CC buffered solution which is pref. a buffered balanced salt solution,
CC at least 1 of hydroxy- or carboxypropyl methylcellulose, a cellulose
CC gum, dextran or dextran sulphate, chondroitin sulphate, and sodium
CC hyaluronate. It has a pH of 6-8 and an osmolality of 200-400 mOsmol/l.
CC The growth factors are mitogenic in vitro for a wide range of tissues
CC and the viscoelastic soln. may be used as a surgical soln. which is
CC in direct contact with cells undergoing wound healing. It also provides
CC a cell protection and cell coating during surgery. The soln. provides
CC maintenance of tissue space, hydroxypropyl methylcellulose and
CC chondroitin sulphate lubricate the tissue, while sodium hyaluronate
CC provides tissue manipulation.

XX Sequence 53 AA;

Query Match 100.0%; Score 54; DB 16; Length 53;

Best Local Similarity 90.0%; Pred. No. 0.33;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10

Db 33 CVIGYSGDRC 42

RESULT 7

AAW34466

ID AAW34466 standard; Protein; 53 AA.

XX AAW34466;

XX 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein AC.

XX Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

XX Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

XX Key Location/Qualifiers

XX Region 1..10

XX /note= "region A from mouse EGF"

XX Region 11..33

XX /note= "region B from human EGF"

XX Region 34..47

XX /note= "region C from mouse EGF"

XX Region 48..53

XX /note= "region D from human EGF"

XX AU9717759-A.

XX 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.

XX

PR 23-APR-1996; 96JP-0123970.

XX (HGET) HIGETA SHOYU KK.

XX Tagami H, Tanaka A;

XX WPI; 1997-550187/51.

XX N-PSDB; AAT99941.

XX Chimeric epidermal growth factor proteins - and DNA molecules for
PT their recombinant production

XX Claim 5; Page 26; 40pp; English.

XX This is a chimeric epidermal growth factor (EGF) protein AC. This
CC chimeric EGF protein contains sequences derived from human and mouse
CC EGF sequences. The chimeric EGF proteins can be recombinantly produced
CC by culturing a microorganism of the genus Bacillus transformed by an
CC expression vector containing the encoding DNA molecules. The recombinant
CC chimeric EGF proteins exhibit a wide variety of physiological activities
CC similar to those of naturally occurring EGF.

XX Sequence 53 AA;

Query Match 100.0%; Score 54; DB 18; Length 53;

Best Local Similarity 90.0%; Pred. No. 0.33;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10

Db 33 CVIGYSGDRC 42

RESULT 8

AAW34467

ID AAW34467 standard; Protein; 53 AA.

XX AAW34467;

XX 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein ABC.

XX Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

XX Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

XX Key Location/Qualifiers

XX Region 1..10

XX /note= "region A from mouse EGF"

XX Region 11..33

XX /note= "region B from mouse EGF"

XX Region 34..47

XX /note= "region C from mouse EGF"

XX Region 48..53

XX /note= "region D from human EGF"

XX AU9717759-A.

XX 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.

XX 23-APR-1996; 96JP-0123970.

XX (HGET) HIGETA SHOYU KK.

XX Tagami H, Tanaka A;

XX WPI; 1997-550187/51.

XX N-PSDB; AAT99942.

XX

PT Chimeric epidermal growth factor proteins - and DNA molecules for
 PT their recombinant production
 XX
 PS Claim 6; Page 26; 40pp; English.
 XX This is a chimeric epidermal growth factor (EGF) protein ABC. This
 CC chimeric EGF protein contains sequences derived from human and mouse
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced
 CC by culturing a microorganism of the genus *Bacillus* transformed by an
 CC expression vector containing the encoding DNA molecules. The recombinant
 CC chimeric EGF proteins exhibit a wide variety of physiological activities
 CC similar to those of naturally occurring EGF.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 54; DB 18; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDRC 10
 Db 33 CVIGYSGDRC 42
 RESULT 9
 AAW50134
 ID AAW50134 standard; protein; 53 AA.
 XX
 AC AAW50134;
 XX
 XX 08-JUL-1998 (first entry)
 DT
 DE Mouse epidermal growth factor.
 XX
 KW Mouse; epidermal growth factor; EGF; isolation; recombinant;
 KW ion exchange chromatography; *Bacillus* brevis.
 XX
 OS Mus sp.
 XX
 PN AU9728698-A.
 XX
 XX 05-FEB-1998.
 PD
 PF 17-JUL-1997; 97AU-0028698.
 XX
 PR 01-AUG-1996; 96JP-0218109.
 XX
 PA (HGET) HIGETA SHOYU KK.
 XX
 PI Miyauchi A, Nemoto A, Warren B;
 XX
 DR WPI; 1998-131057/13.
 XX
 PT Isolation of recombinant epidermal growth factor from whole broth -
 PT by fluidised-bed ion-exchange chromatography
 XX
 PS Claim 6; Page 26; 38pp; English.
 XX
 CC The present sequence represents mouse epidermal growth factor (EGF)
 CC from the present invention. The present invention describes a method for
 CC the isolation of recombinant epidermal growth factor (EGF) from whole
 CC broth. The method comprises: (a) passing a culture containing the
 CC protein upwards through a column containing a fluidised bed of ion
 CC exchanger, washing the ion exchanger to remove non-adsorbed material,
 CC allowing the ion exchanger to settle, and eluting adsorbed protein by
 CC passing an eluant downwards through the column; (b) filtering the eluate
 CC through a membrane with a molecular weight cutoff of 10000, and (c)
 CC concentrating the filtrate on a membrane with a molecular weight cutoff
 CC of 5000. The method is for isolating recombinant human, mouse, pig or
 CC rat EGF, or an EGF analogue with one of four defined amino acid sequences
 CC (given in the specification), from *Bacillus* brevis cultures. The EGF
 CC proteins can be recovered in high yield and high purity without the need
 CC to pretreat the culture broth to remove the cells, e.g. by costly

CC centrifugation or membrane filtration.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 54; DB 19; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDRC 10
 Db 33 CVIGYSGDRC 42
 RESULT 10
 AAW50139
 ID AAW50139 standard; protein; 53 AA.
 XX
 AC AAW50139;
 XX
 XX 08-JUL-1998 (first entry)
 DT
 DE Epidermal growth factor analogue 3.
 XX
 KW Epidermal growth factor; EGF; isolation; recombinant;
 KW ion exchange chromatography; *Bacillus* brevis.
 XX
 OS Unidentified.
 XX
 PN AU9728698-A.
 XX
 XX 05-FEB-1998.
 PD
 PF 17-JUL-1997; 97AU-0028698.
 XX
 PR 01-AUG-1996; 96JP-0218109.
 XX
 PA (HGET) HIGETA SHOYU KK.
 XX
 PI Miyauchi A, Nemoto A, Warren B;
 XX
 DR WPI; 1998-131057/13.
 XX
 PT Isolation of recombinant epidermal growth factor from whole broth -
 PT by fluidised-bed ion-exchange chromatography
 XX
 PS Claim 6; Page 28; 38pp; English.
 XX
 CC The present sequence represents a protein which has epidermal growth
 CC factor (EGF) like activity, from the present invention. The present
 CC invention describes a method for the isolation of recombinant epidermal
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing
 CC a culture containing the protein upwards through a column containing a
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-
 CC adsorbed material, allowing the ion exchanger to settle, and eluting
 CC adsorbed protein by passing an eluant downwards through the column; (b)
 CC filtering the eluate through a membrane with a molecular weight cutoff
 CC of 10000, and (c) concentrating the filtrate on a membrane with a
 CC molecular weight cutoff of 5000. The method is for isolating recombinant
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four
 CC defined amino acid sequences (given in the specification), from *Bacillus*
 CC brevis cultures. The EGF proteins can be recovered in high yield and
 CC high purity without the need to pretreat the culture broth to remove the
 CC cells, e.g. by costly centrifugation or membrane filtration.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 54; DB 19; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDRC 10
 Db 33 CVIGYSGDRC 42

XX PF 21-MAY-2001; 2001WO-GB02226.
 XX PR 19-MAY-2000; 2000GB-0011981.
 XX PR 24-AUG-2000; 2000GB-0020794.
 XX PR (SCAN-) SCANCELL LTD.
 XX PI Ellis JRM, Durrant LG;
 XX WI; 2002-062384/08.
 XX DR
 XX PT New humanized form of mouse monoclonal antibody 340 which binds to
 PT epidermal growth factor receptor and inhibits binding of growth factor,
 PT useful for treating colorectal, lung, breast, gastric and ovarian
 PT cancer -
 XX PS Example 5; Fig 12; 53pp; English.
 XX CC The present invention relates to a humanised form of the antibody 340 (a
 CC mouse monoclonal antibody which binds to epidermal growth factor (EGF)
 CC receptor and inhibits binding of EGF), obtainable from the cell line
 CC deposited with the ECACC under accession number 97021428. The humanised
 CC form of the antibody 340 is useful in gene therapy, medicine and in the
 CC manufacture of a medicament for treatment or prophylaxis of cancer. The
 CC invention is useful for treating colorectal, lung, breast, gastric or
 CC ovarian cancers or also for preventing the recurrence of cancer after
 CC initial treatment or surgery. The invention is also useful for enhancing
 CC a protective immune response against cancer by optimised immunisation
 CC schedules. The humanised form of the antibody 340 has reduced
 CC immunogenicity but shows similar binding to cells expressing EGF
 CC receptor, as the original murine antibody and has increased ability to
 CC inhibit the growth of EGF receptor expressing cells. The invention is
 CC used as cell growth and apoptosis inhibitor. The present sequence
 CC is epidermal growth factor (EGF) which is used in the exemplification of
 CC the invention.
 XX SQ Sequence 53 AA;
 Query Match 100.0%; Score 54; DB 23; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDRC 10
 ||||:||||
 Db 33 CVIGYSGDRC 42
 RESULT 14
 AAU76704
 ID AAU76704 standard; Protein; 54 AA.
 AC AAU76704;
 XX 21-MAY-2002 (first entry)
 DT Mouse epidermal growth factor (EGF).
 DE Mouse; epidermal growth factor; receptor; EGF;
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;
 KW recombination.
 XX Mus sp.
 OS
 XX Key Location/Qualifiers
 FT Misc-difference 54
 FT /label= Unknown
 FT /note= "Encoded by TAA"
 XX WO200206469-A2.
 PN 24-JAN-2002.
 XX

PF 18-JUL-2001; 2001WO-US22640.
 XX 18-JUL-2000; 2000US-218921P.
 PR 18-JUL-2000; 2000US-219085P.
 PR 19-OCT-2000; 2000US-0691873.
 PR 19-OCT-2000; 2000US-0692732.
 XX (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.
 XX Cocco WM, Ensell LP, Arensdorf JJ;
 XX WPI; 2002-179795/23.
 DR N-PSDB; ABK10313.
 XX PT Forming chimeric polynucleotide by contacting single-stranded (ss)
 PT scaffold fragment, and donor fragment populations to form hybridised
 PT complex having ss regions which are filled-in, and ligating adjacent
 PT fragments -
 XX PS Example 2; Fig 3; 65pp; English.
 XX CC The invention describes a method of forming (M1) chimeric
 CC polynucleotides (I). The method comprises contacting a population of
 CC single-stranded scaffold fragments (SF) with population of donor
 CC fragments (DF) to form a complex (II) comprising at least one SF
 CC hybridised to two DFs. (II) is treated so that single stranded regions of
 CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for
 CC a directed evolution process which involves forming a library of (I)
 CC that can be screened for a characteristic of interest. Subsequent rounds
 CC of directed evolution can produce chimeric polynucleotides with an
 CC improved characteristic of interest. The methods facilitate the
 CC generation of chimeric polynucleotides and do not require hybridising
 CC donor fragments to a target- or full-length template. Because the
 CC chimeraogenesis process does not rely upon a contiguous, full-length
 CC template, it is unnecessary to modify a template to facilitate its
 CC removal. This is the amino acid sequence of the mouse epidermal growth
 CC factor (EGF) used to demonstrate a method of in vitro recombination
 CC described in the method of the invention.
 XX SQ Sequence 54 AA;
 Query Match 100.0%; Score 54; DB 23; Length 54;
 Best Local Similarity 90.0%; Pred. No. 0.33;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDRC 10
 ||||:||||
 Db 33 CVIGYSGDRC 42
 RESULT 15
 AAU76706
 ID AAU76706 standard; Protein; 54 AA.
 AC AAU76706;
 XX 21-MAY-2002 (first entry)
 DT Modified mouse epidermal growth factor (EGF).
 DE Mouse; epidermal growth factor; receptor; EGF;
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;
 KW recombination.
 XX Mus sp.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT Misc-difference 38
 FT /note= "Encoded by ACC"
 XX WO200206469-A2.
 PN
 XX

Wed Jul 2 08:08:18 2003

```
PD 24-JAN-2002.
XX
XX PF 18-JUL-2001; 2001WO-US22640.
XX
XX PR 18-JUL-2000; 2000US-218921P.
XX PR 18-JUL-2000; 2000US-219083P.
XX PR 19-OCT-2000; 2000US-0691873.
XX PR 19-OCT-2000; 2000US-0692732.
XX
XX PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.
XX
XX PI Coco WM, Enceill LP, Arensdorf JJ;
XX
XX DR WPI; 2002-179795/23.
XX DR N-PSDB; ABK10315.
XX
XX PT Forming chimeric polynucleotide by contacting single-stranded (ss)
XX PT scaffold fragment, and donor fragment populations to form hybridised
XX PT complex having ss regions which are filled-in, and ligating adjacent
XX PT fragments
XX
XX PS Example 2; Fig 3; 65pp; English.
XX
XX CC The invention describes a method of forming (M1) chimeric
XX CC polynucleotides (I). The method comprises contacting a population of
XX CC single-stranded scaffold fragments (SF) with population of donor
XX CC fragments (DF) to form a complex (II) comprising at least one SF
XX CC hybridised to two Dfs. (II) is treated so that single stranded regions of
XX CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for
XX CC a directed evolution process which involves forming a library of (I)
XX CC that can be screened for a characteristic of interest. Subsequent rounds
XX CC of directed evolution can produce chimeric polynucleotides with an
XX CC improved characteristic of interest. The methods facilitate the
XX CC generation of chimeric polynucleotides and do not require hybridising
XX CC donor fragments to a target- or full-length template. Because the
XX CC chimeragenesis process does not rely upon a contiguous, full-length
XX CC template, it is unnecessary to modify a template to facilitate its
XX CC removal. This is the amino acid sequence of a mouse epidermal growth
XX CC factor (EGF) modified to make it as similar as possible to the human
XX CC protein (AAU76705) and used to demonstrate a method of in vitro
XX CC recombination described in the method of the invention.
XX
XX SQ Sequence 54 AA;
XX
XX Query Match 100.0%; Score 54; DB 23; Length 54;
XX Best Local Similarity 90.0%; Pred. No. 0.33;
XX Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 CVICXSGDRC 10
XX ||||:||||
XX Db 33 CVIGYSGDRC 42
```

Search completed: July 2, 2003, 07:30:54
Job time : 44.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 10.6667 Seconds
(Without alignments)
27.584 Million cell updates/sec

Title: US-09-673-785D-8
Perfect score: 54
Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCRUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	54	100.0	12	1	US-08-597-545-7
2	54	100.0	12	1	US-08-457-135-7
3	54	100.0	53	1	US-08-284-923-2
4	54	100.0	53	4	US-08-619-032B-2
5	54	100.0	53	6	5332669-2
6	44	81.5	48	6	5332669-1
7	44	81.5	754	2	US-08-525-864A-2
8	43	79.6	17	6	5183805-1
9	43	79.6	17	6	5183805-2
10	43	79.6	25	6	5256643-8
11	43	79.6	37	2	US-08-039-364-14
12	43	79.6	37	4	US-08-158-710-14
13	43	79.6	42	1	US-08-168-091A-44
14	43	79.6	44	1	US-08-278-089A-24
15	43	79.6	46	3	US-08-899-437-12
16	43	79.6	46	4	US-09-126-121-12
17	43	79.6	47	3	US-08-753-007A-17
18	43	79.6	47	4	US-09-398-496-17
19	43	79.6	48	4	US-09-020-880-15
20	43	79.6	48	4	US-08-915-096A-13
21	43	79.6	48	4	US-09-101-544-15
22	43	79.6	48	6	5434135-3
23	43	79.6	51	6	517197-50
24	43	79.6	53	1	US-07-869-176-1
25	43	79.6	53	1	US-08-284-923-1
26	43	79.6	53	1	US-08-360-841-2
27	43	79.6	53	2	US-08-861-000-1

28 43 79.6 53 4 US-08-619-032B-1
29 43 79.6 53 6 5332669-3
30 43 79.6 53 6 5434135-2
31 43 79.6 55 6 5218093-1
32 43 79.6 88 2 US-07-885-089B-13
33 43 79.6 91 1 US-07-847-743B-15
34 43 79.6 91 1 US-08-456-201-15
35 43 79.6 91 2 US-08-330-161-13
36 43 79.6 91 2 US-08-456-241-15
37 43 79.6 91 2 US-08-440-401-13
38 43 79.6 91 2 US-08-419-878B-13
39 43 79.6 91 4 US-09-173-480-13
40 43 79.6 91 5 PCT-US92-04295A-15
41 43 79.6 93 1 US-08-343-401A-4
42 43 79.6 93 1 US-08-445-265A-2
43 43 79.6 93 3 US-08-990-442-2
44 43 79.6 132 3 US-08-468-846-13
45 43 79.6 293 4 US-08-438-745-4

ALIGNMENTS

RESULT 1
US-08-597-545-7
; Sequence 7, Application US/08597545
; Patent No. 5580738
; GENERAL INFORMATION:
; APPLICANT: LABORDA, Jorge
; TITLE OF INVENTION: Delta-Like Gene Expressed In
; TITLE OF INVENTION: Neuroendocrine Tumors
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Hardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,545
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/989,537
; FILING DATE: 11-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/166 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-597-545-7

Query Match 100.0%; Score 54; DB 1; Length 12;
Best Local Similarity 90.0%; Pred. No. 0.014;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CVIGXSGDRC 10
Db 3 CVIGXSGDRC 12

RESULT 2

US-08-457-135-7
; Sequence 7, Application US/08457135
; Patent No. 5644031
; GENERAL INFORMATION:
; APPLICANT: LABORDA, Jorge
; TITLE OF INVENTION: Delta-Like Gene Expressed In
; TITLE OF INVENTION: Neuroendocrine Tumors
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,135
; FILING DATE: 01-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/989,537
; FILING DATE: 11-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/304/NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-457-135-7

Query Match 100.0%; Score 54; DB 1; Length 12;
Best Local Similarity 90.0%; Pred. No. 0.014;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 3 CVIGYSGDRC 12

RESULT 3

US-08-284-923-2
; Sequence 2, Application US/08284923
; Patent No. 5547935
; GENERAL INFORMATION:
; APPLICANT: Mullenbach, Guy T
; APPLICANT: Blaney, Jeffrey M
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: Mutins of Epidermal Growth Factor
; TITLE OF INVENTION: exhibiting enhanced binding at low ph
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,923
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/746,651
; FILING DATE: 16-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 231.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-284-923-2

Query Match 100.0%; Score 54; DB 1; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.056;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 33 CVIGYSGDRC 42

RESULT 4

US-08-619-032B-2
; Sequence 2, Application US/08619032B
; Patent No. 6191106
; GENERAL INFORMATION:
; APPLICANT: Mullenbach, Guy T.
; APPLICANT: Blaney, Jeffrey M.
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: MUTINS OF EPIDERMAL GROWTH FACTOR
; TITLE OF INVENTION: EXHIBITING ENHANCED BINDING AT LOW PH
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: Intellectual Property R-440, P.O. Box 8097
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,032B
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Guth, Joseph H.
; REGISTRATION NUMBER: 31,261
; REFERENCE/DOCKET NUMBER: 0231.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3888
; TELEFAX: (510) 655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-619-032B-2

Query Match 100.0%; Score 54; DB 4; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.056;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
| | | | | | | | | |
DB 33 CVIGYSGDRC 42

RESULT 5

5332669-2
; Patent No. 5332669
; APPLICANT: DEUEL, THOMAS F.
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/356.739
; FILING DATE: 24-MAR-1989
; SEQ ID NO:2;
; LENGTH: 53
5332669-2

Query Match 100.0%; Score 54; DB 6; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.056;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
| | | | | | | | | |
DB 33 CVIGYSGDRC 42

RESULT 6

5332669-1
; Patent No. 5332669
; APPLICANT: DEUEL, THOMAS F.
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/356.739
; FILING DATE: 24-MAR-1989
; SEQ ID NO:1;
; LENGTH: 48
5332669-1

Query Match 81.5%; Score 44; DB 6; Length 48;
Best Local Similarity 70.0%; Pred. No. 1.8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
| | | | | | | | | |
DB 33 CVIGYIGERC 42

RESULT 7

US-08-525-864A-2
; Sequence 2, Application US/08525864A
; Patent No. 5912326
; GENERAL INFORMATION:
; APPLICANT: Chang, Han
; TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses
; TITLE OF INVENTION: Related thereto
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,864A
; FILING DATE: 8-SEP-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kara, Catherine J.
; REGISTRATION NUMBER: 41,106
; REFERENCE/DOCKET NUMBER: HUI-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-864A-2

Query Match 81.5%; Score 44; DB 2; Length 754;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
| | | | | | | | | |
DB 280 CPVGYTGDC 289

RESULT 8

5183805-1
; Patent No. 5183805
; APPLICANT: LEE, JIN S.; BLICK, MARK
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/567,407
; FILING DATE: 13-AUG-1990
; SEQ ID NO:1;
; LENGTH: 17
5183805-1

Query Match 79.6%; Score 43; DB 6; Length 17;
Best Local Similarity 60.0%; Pred. No. 0.94;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
| | | | | | | | | |
DB 2 CVVGYIGERC 11

RESULT 9

5183805-2
; Patent No. 5183805
; APPLICANT: LEE, JIN S.; BLICK, MARK
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/567,407
; FILING DATE: 13-AUG-1990
; SEQ ID NO:2;
; LENGTH: 17
5183805-2

Query Match 79.6%; Score 43; DB 6; Length 17;
Best Local Similarity 60.0%; Pred. No. 0.94;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|:|:|:|:|
Db 2 CVVGYIGERC 11

RESULT 10
5256643-8
; Patent No. 5256643
; APPLICANT: Persico, Maria G.; Salomon, David S.
; TITLE OF INVENTION: HUMAN CRYPTO PROTEIN
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/530,165
; FILING DATE: 29-MAY-1990
; SEQ ID NO: 8
; LENGTH: 25
5256643-8

Query Match 79.6%; Score 43; DB 6; Length 25;
Best Local Similarity 60.0%; Pred. No. 1.4;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|:|:|:|:|
Db 13 CVVGYIGERC 22

RESULT 11
US-08-039-364-14
; Sequence 14, Application US/08039364
; Patent No. 5811393
; GENERAL INFORMATION:
; APPLICANT: Klagsbrun, Michael
; APPLICANT: Abraham, Judith A.
; APPLICANT: Higashiyama, Shigeki
; APPLICANT: Besner, Gail F.
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/039,364
; FILING DATE: 15 JUN 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/598,082
; FILING DATE: 16 OCT 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05162/002002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
US-08-039-364-14

Query Match 79.6%; Score 43; DB 2; Length 37;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|:|:|:|:|
Db 28 CVVGYIGERC 37

RESULT 12
US-09-158-710-14
; Sequence 14, Application US/09158710
; Patent No. 6235884
; GENERAL INFORMATION:
; APPLICANT: Klagsbrun, Michael
; APPLICANT: Abraham, Judith A.
; APPLICANT: Higashiyama, Shigeki
; APPLICANT: Besner, Gail F.
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH HOMOLOGY TO EPIDERMAL
; FILE REFERENCE: 05162/002003
; CURRENT APPLICATION NUMBER: US/09/158,710
; CURRENT FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: US 08/039,364
; EARLIER FILING DATE: 1993-06-15
; EARLIER APPLICATION NUMBER: US 07/598,082
; EARLIER FILING DATE: 1990-10-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-158-710-14

Query Match 79.6%; Score 43; DB 4; Length 37;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|:|:|:|:|
Db 28 CVVGYIGERC 37

RESULT 13
US-08-168-091A-44
; Sequence 44, Application US/08168091A
; Patent No. 5665862
; GENERAL INFORMATION:
; APPLICANT: Fischbach, Gerald
; APPLICANT: Falls, Douglas R.
; APPLICANT: Rosen, Kenneth M.
; APPLICANT: Corfas, Gabriel
; TITLE OF INVENTION: Neurotrophic Factor
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE AND COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/168,091A
; FILING DATE: 15-DEC-1993

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/953,742
FILING DATE: 29-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Deconti, Giulio A
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: HMI-002CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-168-091A-44

Query Match 79.6%; Score 43; DB 1; Length 42;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|||:|:|
Db 31 CVVGIGERC 40

RESULT 14
US-08-278-089A-24
Sequence 24, Application US/08278089A
Patent No. 5681714
GENERAL INFORMATION:
APPLICANT: Breitman, Martin L.
APPLICANT: Rossant, Janet
APPLICANT: Dumont, Daniel J.
APPLICANT: Yamaguchi, Terry P.
TITLE OF INVENTION: No. 5681714el Receptor Tyrosine Kinase
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/278,089A
FILING DATE: 20-JUL-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kurdydyk, Linda M.
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 3153-111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 44 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
CLONE: EGF

US-08-278-089A-24

Query Match 79.6%; Score 43; DB 1; Length 44;
Best Local Similarity 60.0%; Pred. No. 2.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|||:|:|
Db 33 CVVGIGERC 42

RESULT 15
US-08-899-437-12
Sequence 12, Application US/08899437
Patent No. 6121415
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
Ligands and Uses Therefor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,437
FILING DATE: 24-Jul-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
FEATURE:
NAME/KEY: hEGF.egf
LOCATION: 1-46
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-899-437-12

Query Match 79.6%; Score 43; DB 3; Length 46;
Best Local Similarity 60.0%; Pred. No. 2.4;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|||:|:|
Db 32 CVVGIGERC 41

Search completed: July 2, 2003, 07:25:43
Job time : 12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:28:49 ; Search time 103 Seconds
(without alignments)
11.164 Million cell updates/sec

Title: US-09-673-785d-8
Perfect score: 54
Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/FCI_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	54	100.0	53	9	US-10-138-158-19
2	54	100.0	53	9	US-10-150-648B-33
3	54	100.0	145	9	US-10-150-648B-30
4	44	81.5	112	9	US-10-211-994-20
5	44	81.5	298	10	US-09-864-675-4
6	44	81.5	878	9	US-10-157-031-171
7	43	79.6	46	9	US-10-136-573A-12
8	43	79.6	46	9	US-09-877-665-12
9	43	79.6	46	9	US-10-215-862-12
10	43	79.6	46	10	US-09-817-647-12
11	43	79.6	47	12	US-10-096-241-17
12	43	79.6	48	9	US-10-201-945-13
13	43	79.6	53	9	US-09-903-327A-9
14	43	79.6	53	9	US-10-150-648B-35
15	43	79.6	53	9	US-10-211-994-4
16	43	79.6	53	10	US-09-848-664-31
17	43	79.6	58	10	US-09-934-706-3
18	43	79.6	91	9	US-10-022-609-13
19	43	79.6	111	9	US-10-211-994-27

20	43	79.6	111	9	US-10-211-994-29	Sequence 29, Appl
21	43	79.6	112	9	US-10-211-994-5	Sequence 5, Appl
22	43	79.6	112	9	US-10-211-994-7	Sequence 7, Appl
23	43	79.6	112	9	US-10-211-994-25	Sequence 25, Appl
24	43	79.6	140	10	US-09-280-030-64	Sequence 64, Appl
25	43	79.6	146	9	US-10-150-648B-32	Sequence 32, Appl
26	43	79.6	400	10	US-09-934-706-5	Sequence 5, Appl
27	43	79.6	493	9	US-09-903-327A-13	Sequence 13, Appl
28	43	79.6	741	10	US-09-925-301-930	Sequence 930, Appl
29	41	75.9	572	9	US-09-900-449A-7	Sequence 7, Appl
30	41	75.9	601	9	US-09-900-449A-5	Sequence 5, Appl
31	41	75.9	639	9	US-09-900-449A-4	Sequence 4, Appl
32	41	74.1	53	9	US-10-150-648B-34	Sequence 34, Appl
33	40	74.1	145	9	US-10-150-648B-31	Sequence 31, Appl
34	40	74.1	176	9	US-10-024-599-4	Sequence 4, Appl
35	40	74.1	1050	9	US-09-796-753-114	Sequence 114, Appl
36	39	72.2	94	10	US-09-764-853-716	Sequence 716, Appl
37	39	72.2	769	9	US-09-984-130-67	Sequence 67, Appl
38	39	72.2	769	9	US-10-097-340-157	Sequence 157, Appl
39	38	70.4	636	9	US-09-796-753-100	Sequence 100, Appl
40	38	70.4	636	9	US-09-796-753-124	Sequence 124, Appl
41	38	70.4	723	10	US-09-893-737-100	Sequence 100, Appl
42	38	70.4	3635	9	US-10-037-182-4	Sequence 4, Appl
43	38	70.4	3635	10	US-09-845-583-2	Sequence 2, Appl
44	38	70.4	3712	9	US-10-108-605-103	Sequence 103, Appl
45	38	70.4	4545	10	US-09-873-403-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-10-138-158-19
; Sequence 19, Application US/10138158
; Publication No. US20030036509A1
; GENERAL INFORMATION:
; APPLICANT: STEM CELL PHARMACEUTICALS, INC.
; APPLICANT: TWARDZIK, Daniel R.
; APPLICANT: PERNET, Andre
; APPLICANT: FELKER, Thomas S.
; APPLICANT: PASKELL, Stefan
; APPLICANT: RENO, John M.
; TITLE OF INVENTION: TGF-alpha POLYPEPTIDES, FUNCTIONAL FRAGMENTS AND METHODS OF
; FILE REFERENCE: STEM110-6
; CURRENT APPLICATION NUMBER: US/10/138.158
; CURRENT FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: US 09/641,587
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/559,248
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/459,813
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 09/492,935
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 09/378,567
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-158-19

Query Match 100.0%; Score 54; DB 9; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.063; Indels 0;
Matches 9; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CVIGXSGDRC 10
| | | | | | | | | |
Db 33 CVIGXSGDRC 42

DATE: 1983
DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
DATABASE ENTRY DATE: 1993-04-27
RELEVANT RESIDUES: Relevant residues FROM 919 TO 1063
US-10-150-648B-30

Query Match	100.0%	Score 54;	DB 9;	Length 145;
Best Local Similarity	90.0%;	Pred NO. 0.16;		
Matches	9;	Conservative	1;	Mismatches 0;
				Indels 0;
QY	1	CVIGXSGDRC	10	
		:		
db	91	CVIGYSGDRC	100	

```

RESULT 4
US-10-211-994-20
; sequence 20, Application US/10211994
; Publication NO. US20030082201A1
; GENERAL INFORMATION:
; APPLICANT: Rao, M.R.S.
; APPLICANT: Sengupta, Paromita
; APPLICANT: Prasad, Sudhanand
; APPLICANT: Burman, Anand C.
; APPLICANT: Mukherjee, Rama
; APPLICANT: Thomas, Becky
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER
; FILE REFERENCE: U014152-1
; CURRENT APPLICATION NUMBER: US/10/211,994
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/309,975
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide and peptide sequence of multivalent vaccine
US-10-211-994-20

```

Query Match 81.5%; Score 44; DB 9; Length 112;
Best Local Similarity 70.0%; Pred. No. 5.1;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY	1	CVIGXSGDRC	10
		:	
ph	92	CVIGYIGERC	101

```

RESULT 5
US-09-864-675-4
  ? Sequence 4, Application US/09864675
  ? Patent No. US20020081286A1
  ? GENERAL INFORMATION:
  ? APPLICANT: Marchionni, Mark
  ? TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,
  ? POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS
  ? FILE REFERENCE: 04595/049002
  ? CURRENT APPLICATION NUMBER: US/09/864,675
  ? CURRENT FILING DATE: 2001-05-23
  ? PRIOR APPLICATION NUMBER: 495
  ? PRIOR FILING DATE: 2000-05-23
  ? NUMBER OF SEQ ID NOS: 18
  ? SOFTWARE: FastSeq for Windows Version 4.0
  ? SEQ ID NO 4
  ? LENGTH: 298
  ? TYPE: PRT
  ? ORGANISM: Homo sapiens
  ? US-09-864-675-4

```

Query Match
81.5%; Score 44; DB 10; Length 298;

```

RESEQUENCE 2
US-10-150-648B-33
; Sequence 33, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Billodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Sellinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725
; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 977 TO 1029
US-10-150-648B-33

```

Query Match 100.0%; Score 54; DB 9; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.063;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 CVIGXSGDRC 10
 ||||:|||||
 33 CVIGYSGDRC 42

```

RESULT 3
US-10-150-648B-30
; Sequence 30, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725

```

```
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 280 CPVGTGDC 289

RESULT 6
US-10-157-031-171
; Sequence 171, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yanovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 171
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-171

Query Match 81.5%; Score 44; DB 9; Length 878;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 526 CLPFGSGDC 535

RESULT 7
US-10-136-573A-12
; Sequence 12, Application US/10136573A
; Patent No. US20020161200A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; FILE REFERENCE: P1084RIC2
; CURRENT APPLICATION NUMBER: US/10/136,573A
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: US 09/480,977
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 08/899,437
; PRIOR FILING DATE: 1997-07-24
; PRIOR APPLICATION NUMBER: US 60/052,019
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 12
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-136-573A-12

Query Match 79.6%; Score 43; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 32 CVVGYIGRC 41

RESULT 8
US-09-877-665-12
; Sequence 12, Application US/09877665
; Patent No. US20020164680A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/877,665
; FILING DATE: 08-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/109,206
; FILING DATE: 30-Jun-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; TYPE: Amino Acid
; LENGTH: 46 amino acids
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: hEGF.egf
; LOCATION: 1-46
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-877-665-12

Query Match 79.6%; Score 43; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 32 CVVGYIGRC 41

RESULT 9
US-10-215-862-12
; Sequence 12, Application US/10215862
; Publication No. US20030036166A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; FILE REFERENCE: P1084R1D2C1
; CURRENT APPLICATION NUMBER: US/10/215,862
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 09/126,663
; PRIOR FILING DATE: 1998-07-30
```

;; PRIOR APPLICATION NUMBER: US 08/899,437
;; PRIOR FILING DATE: 1997-07-24
;; PRIOR APPLICATION NUMBER: US 60/052,019
;; PRIOR FILING DATE: 1997-07-09
;; NUMBER OF SEQ ID NOS: 23
;; SEQ ID NO 12
;; LENGTH: 46
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-215-862-12

Query Match 79.6%; Score 43; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 32 CVVGYIGRC 41
||:|: |||

RESULT 10
US-09-817-647-12
; Sequence 12, Application US/09817647
; Patent No. US20020082229A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/817,647
FILING DATE: 26-Mar-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/107,979
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Conley, Delidre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084r1-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
FEATURE:

NAME/KEY: hEGF.egf
LOCATION: 1-46
IDENTIFICATION METHOD:
OTHER INFORMATION:
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-817-647-12

Query Match 79.6%; Score 43; DB 10; Length 46;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

Db 32 CVVGYIGRC 41
||:|: |||

RESULT 11
US-10-096-241-17
; Sequence 17, Application US/10096241
; Patent No. US20020127594A1
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; Busfield, Samantha J.

; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/096,241
FILING DATE: 12-Mar-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,591
FILING DATE: 19-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter

REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07334/022001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-096-241-17

Query Match 79.6%; Score 43; DB 12; Length 47;
Best Local Similarity 60.0%; Pred. No. 3.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 33 CVVGYIGRC 42
||:|: |||

RESULT 12
US-10-201-945-13
; Sequence 13, Application US/10201945
; Publication No. US20020188110A1
; GENERAL INFORMATION:
; APPLICANT: Meissner, Paul S.
; Fuldner, Rebecca A.
; Adams, Mark D.

; TITLE OF INVENTION: Transforming Growth Factor Alpha HI
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue

CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/201,945
FILING DATE: 25-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/471,377
FILING DATE: 23-Dec-1999
APPLICATION NUMBER: 08/915,096
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/208,008
FILING DATE: 08-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF110D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEX: 301-309-8439
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-201-945-13
Query Match 79.6%; Score 43; DB 9; Length 48;
Best Local Similarity 60.0%; Pred. No. 3.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
DB 28 CVVGYIGERC 37
RESULT 13
US-09-903-327A-9
Sequence 9, Application US/09903327A
Patent No. US20020164333A1
GENERAL INFORMATION:
APPLICANT: Nemerow, Glen R.
TITLE OF INVENTION: BIFUNCTIONAL MOLECULES AND VECTORS COMPLEXED THEREWITH FOR TARGET
TITLE OF INVENTION: GENE
TITLE OF INVENTION: DELIVERY
FILE REFERENCE: 22908-1228
CURRENT APPLICATION NUMBER: US/09/903,327A
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 09/613,017
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 53
TYPE: PRT
ORGANISM: Human
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (0)...(0)
OTHER INFORMATION: Epidermal Growth Factor (EGF, mature peptide)
US-09-903-327A-9

Query Match 79.6%; Score 43; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. No. 3.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
DB 33 CVVGYIGERC 42
RESULT 14
US-10-150-648B-35
Sequence 35, Application US/10150648B
Publication No. US20030059802A1
GENERAL INFORMATION:
APPLICANT: Bilodeau-Goesseels, Sylvie
APPLICANT: John, Sushil J.
APPLICANT: Selinger, Leonard B..
APPLICANT: Benkel, Bernhard F..
TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
TITLE OF INVENTION: factor
FILE REFERENCE: 60-01
CURRENT APPLICATION NUMBER: US/10/150,648B
CURRENT FILING DATE: 2002-05-17
PRIOR APPLICATION NUMBER: 60/292,136
PRIOR FILING DATE: 2001-05-18
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35
LENGTH: 53
TYPE: PRT
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
AUTHORS: Bell, G.I., Fong, N.M., Stempieu, M.M., Wormsted, M.A.,
AUTHORS: Caput, D., Ku, L.L., Urdea, M.S., Rall, L.B. and Sanchez-
AUTHORS: Pescador, R..
TITLE: Human epidermal growth factor precursor: cDNA sequence,
TITLE: expression in vitro and gene organization.
JOURNAL: Nucleic Acids Research
VOLUME: 14
ISSUE: 21
PAGES: 8427-8446
DATE: 1986
DATABASE ENTRY DATE: 1993-04-21
DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 X04571
RELEVANT RESIDUES: Relevant residues FROM 970 TO 1022
US-10-150-648B-35
Query Match 79.6%; Score 43; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. No. 3.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
DB 33 CVVGYIGERC 42
RESULT 15
US-10-211-994-4
Sequence 4, Application US/10211994
Publication No. US20030082201A1
GENERAL INFORMATION:
APPLICANT: Rao, M.R.S.
APPLICANT: Sengupta, Paromita
APPLICANT: Prasad, Sudhanand
APPLICANT: Burman, Anand C..
APPLICANT: Mukherjee, Rama
APPLICANT: Thomas, Becky
TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER
FILE REFERENCE: U014152-1
CURRENT APPLICATION NUMBER: US/10/211,994
CURRENT FILING DATE: 2002-08-02
PRIOR APPLICATION NUMBER: 60/309,975
PRIOR FILING DATE: 2001-08-03

```
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Epidermal growth factor
US-10-211-994-4

Query Match      79.6%; Score 43; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. No. 3.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CVIGXSGDRC 10
   |||: |||
Db 33 CVVGYIGERC 42

Search completed: July 2, 2003, 07:57:18
Job time : 103 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 12 Seconds
(without alignments)
80.112 Million cell updates/sec

Title: US-09-673-785D-8

Perfect score: 54

Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	1217	1 EGMSMG	epidermal growth f
2	49	90.7	2139	2 A35672	crumbs protein - f
3	47	87.0	1643	2 T14274	versican precursor
4	47	87.0	3381	2 T42389	versican precursor
5	44	81.5	57	2 P04415	ErbB kinase activa
6	44	81.5	372	2 T29359	hypothetical prote
7	44	81.5	1133	1 EGRT	epidermal growth f
8	43	79.6	102	2 B55885	chondroitin sulfat
9	43	79.6	264	2 T16271	hypothetical prote
10	43	79.6	862	2 S43922	versican - pig-tai
11	43	79.6	1207	1 EGHU	epidermal growth f
12	43	79.6	2409	1 EG0979	versican precursor
13	42	77.8	2352	2 T30201	Notch homolog prot
14	42	77.8	2531	2 T31070	notch homolog - se
15	42	77.8	4006	2 T09070	probable tenascin
16	41	75.9	447	2 A39321	mucln - rat (fragm
17	41	75.9	601	2 T48539	probable potassium
18	41	75.9	1064	2 A40136	fibropellin Ia - s
19	40	74.1	53	2 S17294	epidermal growth f
20	40	74.1	57	2 B69300	hypothetical prote
21	40	74.1	601	2 T34396	hypothetical prote
22	40	74.1	1203	2 A49175	Notch B protein -
23	40	74.1	2471	2 A49128	cell-fate determin
24	39	72.2	230	2 A44074	probable EGF-like
25	39	72.2	768	2 B41029	integrin beta-8 ch
26	39	72.2	769	2 A41029	integrin beta-8 ch
27	39	72.2	1111	2 T26972	hypothetical prote
28	39	72.2	1531	2 T42218	slit-1 protein hom
29	39	72.2	1964	2 T09059	notch4 - mouse

ALIGNMENTS

RESULT 1

EGMSMG

epidermal growth factor precursor - mouse

N:Alternate names: urogastrone precursor

C:Species: Mus musculus (house mouse)

C>Date: 30-Nov-1980 #sequence.revision 11-Aug-1983 #text_change 19-Jan-2001

C:Accession: A94272; A93304; A92118; A01387

R:Scott, J.; Urdea, M.; Quiroga, M.; Sanchez-Pescador, R.; Fong, N.; Selby, M.; Ruttenberg, A.

Science 221, 236-240, 1983

A:Title: Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor

A:Reference number: A94272; MUID:83223630; PMID:6602382

A:Accession: A94272

A:Molecule type: mRNA

A:Residues: 1-1217 <SC>

A:Cross-references: GB:J00380; NID:gl92993; PIDN:AAA37539.1; PID:g309210

R:Gray, A.; Dull, T.; Ullrich, A.

Nature 303, 722-725, 1983

A:Title: Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein

A:Reference number: A93304; MUID:83219309; PMID:6304537

A:Accession: A93304

A:Molecule type: mRNA

A:Residues: 1-789,'Y',791-1047,'S',1049-1168 <GRA>

A:Cross-references: GB:J00380

A:Note: the sequence shown by these authors differs from residues 1134-1168 due to an error in the original publication

R:Savage Jr., C.R.; Inagami, T.; Cohen, S.

J. Biol. Chem. 247, 7612-7621, 1972

A:Title: The primary structure of epidermal growth factor.

A:Reference number: A92118; MUID:73048516; PMID:4636327

A:Accession: A92118

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Accession: A92144

A:Contents: annotation; disulfide bonds

C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiation of many cell types

C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in some species) in the kidney

C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound protein

C:Comment: The active growth factor from this submaxillary gland protein stimulates the proliferation of many cell types

C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-containing superfamily

C:Keywords: duplication; growth factor; tandem repeat; transmembrane protein

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1217/Product: epidermal growth factor proprotein

F:29-1038/Domain: extracellular #status predicted <EXT>

F:50-485/Region: EGF precursor long repeat

F:53-92/Domain: LDL receptor YWTD-containing repeat homology <YW01>

F:93-134/Domain: LDL receptor YWTD-containing repeat homology <YW02>

versican precursor
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
laminin alpha-2 ch
laminin alpha 5 ch
hypothetical prote
probable laminin a
laminin alpha-1 ch
alpha-2-macroglobu
-alpha-2-macroglobu
hydrophobin Cg-2
neuroendocrine pro
60S ribosomal prot
hypothetical prote

30 72.2 2397 1 A55535
31 72.2 4307 2 T20721
32 70.4 117 2 T38295
33 71.6 2 T51426
34 1086 2 T05407
35 70.4 3106 1 S53868
36 70.4 3635 2 T10053
37 70.4 3672 2 T23433
38 70.4 3704 2 T37316
39 70.4 3712 2 S18253
40 70.4 4544 1 S02392
41 70.4 4545 1 S25111
42 70.4 108 2 A46222
43 68.5 161 2 S03938
44 68.5 174 2 B90094
45 68.5 422 2 S74581

F:135-176/Domain: LDL receptor YWTD-containing repeat homology <YW03>
F:177-217/Domain: LDL receptor YWTD-containing repeat homology <YW04>
F:218-262/Domain: LDL receptor YWTD-containing repeat homology <YW05>
F:263-307/Domain: LDL receptor YWTD-containing repeat homology <YW06>
F:324-360/Domain: EGF homology <status atypical <EG1>
F:366-401/Domain: EGF homology <EG2>
F:407-442/Domain: EGF homology <EG3>
F:445-482/Domain: EGF homology <EG4>
F:486-529/Region: EGF precursor long repeat
F:489-529/Domain: LDL receptor YWTD-containing repeat homology <YW07>
F:530-572/Domain: LDL receptor YWTD-containing repeat homology <YW08>
F:573-615/Domain: LDL receptor YWTD-containing repeat homology <YW09>
F:616-659/Domain: LDL receptor YWTD-containing repeat homology <YW10>
F:660-700/Domain: LDL receptor YWTD-containing repeat homology <YW11>
F:701-743/Domain: LDL receptor YWTD-containing repeat homology <YW12>
F:751-786/Domain: EGF homology <EG5>
F:842-875/Domain: EGF homology <EG6>
F:881-917/Domain: EGF homology <EG7>
F:923-958/Domain: EGF homology <EG8>
F:977-1029/Product: epidermal growth factor #status experimental <EGF>
F:982-1018/Domain: EGF homology <EG9>
F:1039-1063/Domain: transmembrane #status predicted <TM>
F:1064-1217/Domain: intracellular #status predicted <INT>
F:347-360,366-377,373-386,388-401,407-418,414-427,429-442,445-457,453-467,469-482,751-786
tus predicted
F:982-996,990-1007,1009-1018/Disulfide bonds: #status experimental

Query Match 100.0%; Score 54; DB 1; Length 1217;
Best Local Similarity 90.0%; Pred. No. 0.32;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|||:||||
Db 1009 CVIGXSGDRC 1018

RESULT 2

A35672
crumbs protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 11-Jan-2000
C:Accession: A35672

R:Repass, U.; Theres, C.; Knust, E.
Cell 61, 787-799, 1990

A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila

A:Reference number: A35672; MUID:90263104; PMID:2344615

A:Accession: A35672

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2139 <TEP>

A:Cross-references: GB:M33753

C:Genetics:

A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue

A:Gene: FlyBase:crb

A:Cross-references: FlyBase:FBgn0000368

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: transmembrane protein

F:352-385/Domain: EGF homology <EGX1>

F:392-424/Domain: EGF homology <EGF1>

F:691-722/Domain: EGF homology <EGF>

F:767-799/Domain: EGF homology <EGF3>

F:1878-1914/Domain: EGF homology <EGX2>

Query Match 90.7%; Score 49; DB 2; Length 2139;

Best Local Similarity 70.0%; Pred. No. 3.8;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|:|:|||||

Db 571 CAVGSGDRC 580

RESULT 3

T14274

versican precursor, splice form V2 - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000
C:Accession: T14274

R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.

J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine

A:Reference number: Z17954; MUID:98288320; PMID:9624174

A:Accession: T14274

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1643 <SCH>

A:Cross-references: EMBL:AF060458; NID:g3253303; PID:g3253304; PIDN:AAC24360.1

A:Experimental source: brain

C:Keywords: glycoprotein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-1643/Product: versican, splice form V2 #status predicted <MAT>

F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent)

Query Match 87.0%; Score 47; DB 2; Length 1643;

Best Local Similarity 80.0%; Pred. No. 6.6;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|||:|||||

Db 1362 CVPGYSGDRC 1371

RESULT 4

T42389

versican precursor, splice form V0 - bovine

N:Alternate names: chondroitin sulfate proteoglycan

C:Species: Bos primigenius taurus (cattle)

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000

C:Accession: T42389

R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.

J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine

A:Reference number: Z17954; MUID:98288320; PMID:9624174

A:Accession: T42389

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-3381 <SCH>

A:Cross-references: EMBL:AF060456; NID:g32533299; PID:g3253300; PIDN:AAC24358.1

C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type 1

C:Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-3381/Product: versican, splice form V0 #status predicted <MAT>

F:57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2336

Query Match 87.0%; Score 47; DB 2; Length 3381;

Best Local Similarity 80.0%; Pred. No. 13;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
|||:|||||

Db 3100 CVPGYSGDRC 3109

RESULT 5

PC4415

ErbB kinase activator beta, brain and thymus - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Dec-1997 #sequence_revision 10-Dec-1997 #text_change 02-Aug-2002

C:Accession: PC4415

R:Hisagashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Mi

J. Biochem. 122, 675-680, 1997

A:Title: A novel brain-derived

A:Reference number: JC5700; MUID:98006324; PMID:9348101

A:Accession: PC4415

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-57 <HIG>
A:Cross-references: DDBJ:D89997; NID:G2605633; PIDN:BAA23346.1; PID:G2605634
A:Experimental source: PC-12 cell
C:Comment: This protein is a member of the epidermal growth factor family. It is functional in the differentiation of MDA-MB-453 cells.
C:Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology F.1-25/Domain: EGF homology (fragment) <EGF>
Query Match 81.5%; Score 44; DB 2; Length 57;
Best Local Similarity 60.0%; Pred. No. 1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
| :|||:
Db 16 CPVGTGDRG 25
| :|||:
RESULT 6
T29359
hypothetical protein R05G6.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29359
R:Murray, J.; Ie, T.T.
A:Description: The sequence of C. elegans cosmid R05G6.
A:Reference number: Z20612
A:Accession: T29359
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-372 <MUR>
A:Cross-references: EMBL:U58746; PIDN:AAB00626.1; GSPDB:GN00022; CESP:R05G6.9
A:Experimental source: strain Bristol N2; clone R05G6
C:Genetics:
A:Gene: CESP:R05G6.9
A:Map position: 4
A:Intron: 80/1; 161/1; 245/3; 286/1
Query Match 81.5%; Score 44; DB 2; Length 372;
Best Local Similarity 60.0%; Pred. No. 5.6;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
| :|||:
Db 233 CVLGYSGDRC 242
| :|||:
RESULT 7
EGF
epidermal growth factor precursor - rat
N:Alternate names: urogastrone precursor
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 14-Aug-1998 #text_change 18-Jun-1999
C:Accession: I52995; S05074; S01974; A25425; S18419; S08288
R:Sagui, S.J.; Safirstein, R.; Price, P.M.
DNA Cell Biol. 11, 481-487, 1992
A:Title: Cloning and Sequencing of the Rat Preproepidermal Growth Factor cDNA: Comparison of the Rat and Human Sequences
A:Reference number: I52995; MUID:92398779; PMID:1524680
A:Accession: I52995
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1133 <RES>
A:Cross-references: EMBL:U04842; NID:G440236; PIDN:AAB60436.1; PID:G440237
R:Simpson, R.
submitted to the EMBL Data Library, August 1988
A:Reference number: S05074
A:Accession: S05074
A:Molecule type: mRNA
A:Residues: 'W', 966, 'RWL', 970-1023, 'NW', 1026-1108, 'SGAGVSSGPPQWFWVLE', 1126, 'HQ' <SIM>
A:Cross-references: EMBL:X12748
R:Dorow, D.S.; Simpson, R.J.
Nucleic Acids Res. 16, 9338, 1988
A:Title: Cloning and sequence analysis of a cDNA for rat epidermal growth factor.

A:Reference number: S01974; MUID:89016634; PMID:3262867
A:Accession: S01974
A:Molecule type: mRNA
A:Residues: 'W', 966, 'RWL', 970-1023, 'NW', 1026-1108 <DOR>
A:Cross-references: EMBL:X12748
R:Simpson, R.J.; Smith, J.A.; Moritz, R.L.; O'Hare, M.J.; Rudland, P.S.; Morrison, J.
Eur. J. Biochem. 153, 629-637, 1985
A:Title: Rat epidermal growth factor: complete amino acid sequence.
A:Reference number: A25425; MUID:86081810; PMID:3000782
A:Accession: A25425
A:Molecule type: protein
A:Residues: 974-1021 <SI2>
R:Nishi, N.; Shimizu, C.; Okutani, T.; Kagawa, Y.; Takasuga, H.; Suno, M.; Wada, F.
Biochim. Biophys. Acta 1095, 268-275, 1991
A:Title: Rat prostatic growth factors: purification and characterization of high and low molecular weight forms.
A:Reference number: S18419; MUID:92069070; PMID:1958699
A:Accession: S18419
A>Status: preliminary
A:Molecule type: protein
A:Residues: 974-1021 <NIS>
R:Nexo, E.; Jorgensen, P.E.; Thim, L.; Roepstorff, P.
Biochim. Biophys. Acta 1037, 388-393, 1990
A:Title: Purification and characterization of a low and a high molecular weight form of the epidermal growth factor precursor.
A:Reference number: S08288; MUID:90181442; PMID:2310752
A:Accession: S08288
A:Molecule type: protein
A:Residues: 974-1024 <NEX>
C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiation of epithelial cells.
C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in the case of EGF precursor) in the kidney as a receptor-like membrane-bound protein.
C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor WYTD-CO
C:Keywords: duplication; growth factor; mitogen; tandem repeat; transmembrane protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1133/Product: epidermal growth factor proprotein, membrane-bound form #status predicted <EXT>
F:44-480/Region: EGF precursor long repeat <LR1>
F:47-86/Domain: LDL receptor WYTD-containing repeat homology <YW01>
F:87-128/Domain: LDL receptor WYTD-containing repeat homology <YW02>
F:129-170/Domain: LDL receptor WYTD-containing repeat homology <YW03>
F:171-212/Domain: LDL receptor WYTD-containing repeat homology <YW04>
F:213-257/Domain: LDL receptor WYTD-containing repeat homology <YW05>
F:258-302/Domain: LDL receptor WYTD-containing repeat homology <YW06>
F:319-355/Domain: EGF homology <EG2>
F:361-396/Domain: EGF homology <EG3>
F:402-437/Domain: EGF homology <EG4>
F:440-477/Domain: EGF precursor long repeat <LR2>
F:482-558/Region: EGF precursor long repeat <LR2>
F:485-525/Domain: LDL receptor WYTD-containing repeat homology <YW07>
F:526-568/Domain: LDL receptor WYTD-containing repeat homology <YW08>
F:569-611/Domain: LDL receptor WYTD-containing repeat homology <YW09>
F:612-655/Domain: LDL receptor WYTD-containing repeat homology <YW10>
F:656-696/Domain: LDL receptor WYTD-containing repeat homology <YW11>
F:697-739/Domain: LDL receptor WYTD-containing repeat homology <YW12>
F:747-782/Domain: EGF homology <EG5>
F:839-872/Domain: EGF homology <EG6>
F:878-914/Domain: EGF homology <EG7>
F:920-955/Domain: EGF homology <EG8>
F:974-1024/Product: epidermal growth factor #status experimental <MAT>
F:979-1015/Domain: EGF homology <EG9>
F:1036-1060/Domain: transmembrane #status predicted <TMM>
F:1061-1133/Domain: intracellular #status predicted <INT>
F:342-355, 361-372, 368-381, 383-396, 402-413, 409-422, 424-437, 440-452, 448-462, 464-477, 74-1015/Disulfide bonds: #status predicted

Query Match 81.5%; Score 44; DB 1; Length 1133;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
| :|||:
Db 1006 CVIGYIGERC 1015
| :|||:
| :|||:

RESULT 8

B5885
chondroitin sulfate proteoglycan MV3 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 16-Jul-1999
C:Accession: B5885
R:Zako, M.; Shinomura, T.; Ujita, M.; Ito, K.; Kimata, K.
J. Biol. Chem. 270, 3914-3918, 1995
A:Title: Expression of PG-M(V3), an alternatively spliced form of PG-M without a chondroitin sulfate domain, in chondrocytes of rat articular cartilage.
A:Reference number: A55885; MUID:95181355; PMID:7876137
A:Accession: B5885
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-102 <ZAK>
A:Cross-references: GB:S75879; GB:D32039
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EGF homology <EG1>
F:62-93/Domain: EGF homology <EG2>

Query Match 79.6%; Score 43; DB 2; Length 102;

Best Local Similarity 70.0%; Pred. No. 2.6;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

DB 46 CVPYSGDQC 55

RESULT 9

T16271
hypothetical protein F35D2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16271
R:Connell, M.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F35D2.
A:Reference number: 218488
A:Accession: T16271
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-264 <CON>
A:Cross-references: EMBL:U28741; NID:g861290; PID:g861291; PIDN:AAA68325.1; CESP:F35D2.3
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F35D2.3
A:Introns: 40/3; 71/3; 160/3; 197/3

Query Match 79.6%; Score 43; DB 2; Length 264;

Best Local Similarity 70.0%; Pred. No. 6.2;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

DB 76 CPGYLSGDRC 85

RESULT 10

S43922
versican - pig-tailed macaque (fragments)
N:Alternate names: chondroitin sulfate proteoglycan
C:Species: Macaca nemestrina (pig-tailed macaque)
C>Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 16-Jul-1999
C:Accession: S43922
R:Iso, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.
Matrix Biol. 14, 213-225, 1994
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by immunohistochemistry.
A:Reference number: S43921; MUID:95005762; PMID:7921538
A:Accession: S43922
A:Molecule type: mRNA
A:Residues: 1-233; 234-525; 526-862 <YAO>
A:Cross-references: EMBL:S72413

A:Note: 507-Ser was also found

A:Note: the authors translated the codon GCC for residue 50 as Val, AAG for residue 4

669 as Asn

C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;

F:1-37/Domain: link protein repeat homology (fragment) <LNK1>

F:58-139/Domain: link protein repeat homology (fragment) <LNK1>

F:722-753/Domain: EGF homology <EG1>

F:760-791/Domain: EGF homology <EG2>

Query Match 79.6%; Score 43; DB 2; Length 862;

Best Local Similarity 70.0%; Pred. No. 18;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10

DB 744 CVPYSGDQC 753

RESULT 11

EGH9

epidermal growth factor precursor [validated] - human

N:Alternate names: urogastrone precursor

C:Species: Homo sapiens (man)

C>Date: 30-Nov-1980 #sequence_revision 14-Aug-1998 #text_change 08-Dec-2000

C:Accession: A25531; A01388; A33517; A29721; S45282; S45283

R:Bell, G.I.; Fong, N.M.; Stempfen, M.M.; Wormsted, M.A.; Caput, D.; Ku, L.; Urdea, M.

Nucleic Acids Res. 14, 8427-8446, 1986

A:Title: Human epidermal growth factor precursor: cDNA sequence, expression in vitro

A:Reference number: A25531; MUID:87066721; PMID:3491360

A:Accession: A25531

A:Molecule type: mRNA

A:Residues: 1-1207 <BEL>

A:Cross-references: EMBL:X04571; NID:g31120; PIDN:CAA28240.1; PID:g31121

A:Note: 708-Met was also found

A:Note: intron positions were also determined

R:Gregory, H.; Preston, B.M.

Int. J. Pept. Protein Res. 9, 107-118, 1977

A:Title: The primary structure of human urogastrone.

A:Reference number: A01388; MUID:77117897; PMID:300079

A:Accession: A01388

A:Molecule type: protein

A:Residues: 971-1023 <GRE>

A:Note: some of the molecules lack Arg-1023

R:Furuya, M.; Akashi, S.; Hirayama, K.

Biochem. Biophys. Res. Commun. 163, 1100-1106, 1989

A:Title: The primary structure of human EGF produced by genetic engineering, studied

A:Reference number: A33517; MUID:89391964; PMID:2789514

A:Accession: A33517

A:Molecule type: protein

A:Residues: 971-1023 <FUR>

R:Tsukumo, K.; Nakamura, H.; Sakamoto, S.

Biochem. Biophys. Res. Commun. 145, 126-133, 1987

A:Title: Purification and characterization of high molecular weight human epidermal

A:Reference number: A29721; MUID:87241488; PMID:3297054

A:Accession: A29721

A:Molecule type: protein

A:Residues: 829-834, 'X', 836-839, 'X', 841-845, 'X', 847-848 <TSU>

A:Note: this is the amino-terminal sequence of a high molecular weight form of EGF.

R:Svoboda, M.; Bauhofer, A.; Schwind, P.; Bade, E.; Rasched, I.; Przybylski, M.

Biochim. Biophys. Acta 1206, 35-41, 1994

A:Title: Structural characterization and biological activity of recombinant human epi

A:Reference number: S45282; MUID:94242778; PMID:8186248

A:Accession: S45282

A:Molecule type: protein

A:Residues: 'M', 971-1023 <SVO>

A:Note: expressed recombinant protein

A:Accession: S45283

A:Molecule type: protein

A:Residues: 'MKYP', 970-1023 <SV2>

A:Note: expressed recombinant protein

C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiation

gastrointestinal cell proliferation.

C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in some C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound protein

C:Gene: GDB:EGF
A:Cross-references: GDB:119105; OMIM:131530
A:Map position: 4q25-4q25
A:Introns: 43/1; 109/3; 170/2; 246/2; 314/1; 356/3; 397/1; 438/1; 480/1; 525/3; 575/2; 6
C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-contain
C:Keywords: duplication; growth factor; tandem repeat; transmembrane protein
F:1-22/Domain: signal sequence status predicted <SIG>
F:23-1207/Product: epidermal growth factor proprotein, membrane-bound form #status predi
F:23-1032/Domain: extracellular #status predicted <EXT>
F:43-479/Region: EGF precursor long repeat <LR1>
F:46-85/Domain: LDL receptor YWTD-containing repeat homology <YW01>
F:86-127/Domain: LDL receptor YWTD-containing repeat homology <YW02>
F:128-169/Domain: LDL receptor YWTD-containing repeat homology <YW03>
F:170-211/Domain: LDL receptor YWTD-containing repeat homology <YW04>
F:212-256/Domain: LDL receptor YWTD-containing repeat homology <YW05>
F:257-301/Domain: LDL receptor YWTD-containing repeat homology <YW06>
F:318-354/Domain: LDL receptor YWTD-containing repeat homology <EG1>
F:360-395/Domain: EGF homology <EG2>
F:401-436/Domain: EGF homology <EG3>
F:439-476/Domain: EGF homology <EG4>
F:480-523/Region: EGF precursor long repeat <LR2>
F:483-523/Domain: LDL receptor YWTD-containing repeat homology <YW07>
F:524-566/Domain: LDL receptor YWTD-containing repeat homology <YW08>
F:567-609/Domain: LDL receptor YWTD-containing repeat homology <YW09>
F:610-653/Domain: LDL receptor YWTD-containing repeat homology <YW10>
F:654-694/Domain: LDL receptor YWTD-containing repeat homology <YW11>
F:695-737/Domain: LDL receptor YWTD-containing repeat homology <YW12>
F:745-780/Domain: EGF homology <EG5>
F:835-868/Domain: EGF homology <EG6>
F:874-910/Domain: EGF homology <EG7>
F:916-951/Domain: EGF homology <EG8>
F:971-1023/Product: epidermal growth factor #status experimental <EGF>
F:976-1012/Domain: EGF homology <EG9>
F:1033-1057/Domain: transmembrane #status predicted <TM>
F:1058-1207/Domain: intracellular #status predicted <INT>
F:318-330,325-339,341-354,360-371,367-380,382-395,401-412,408-421,423-436,439-451,447-448
fide bonds: #status predicted
F:976-990,984-1001,1003-1012/Disulfide bonds: #status experimental

Query Match 79.6%; Score 43; DB 1; Length 1207;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
DB 1003 CVVXIGERC 1012

RESULT 12

A60979
versican precursor - human
N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan
N:Contains: glial hyaluronate-binding protein
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Jan-2001
C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179
EMBO J. 8, 2975-2981, 1989
R:Zimmermann, D.R.; Ruoslahti, E.
A:Title: Multiple domains of the large fibroblast proteoglycan, versican.
A:Reference number: S06014; MUID:90059882; PMID:2583089
A:Accession: S06014
A:Molecule type: mRNA
A:Residues: 1-2409 <ZIM>
A:Cross-references: GB:X15998; NID:937662; PIDN:CAA34128.1; PID:g37663
R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.
Matrix Biol. 14, 213-225, 1994
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by
A:Reference number: S43921; MUID:95005762; PMID:7921538
A:Accession: S43921
A:Molecule type: mRNA

A:Residues: 208-440;1094-1385;1910-2246 <YAO>
R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.
Brain Res. Bull. 22, 67-70, 1989
A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilag
A:Reference number: A60979; MUID:89229983; PMID:2469524
A:Accession: A60979
A:Molecule type: protein
A:Residues: 171-210;289-303 <BIG>
R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.
J. Biol. Chem. 264, 5981-5987, 1989
A:Title: Isolation and partial characterization of a glial hyaluronate-binding prote
A:Reference number: A30358; MUID:89174663; PMID:2466833
A:Accession: A30358
A:Molecule type: protein
A:Residues: 24-50;80-87; 'D'; 89-119;128-155;167-218;229-259; 'IR';261-268;277-283; 'G'
R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.
J. Biol. Chem. 262, 13120-13125, 1987
A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-
A:Reference number: A29348; MUID:88007514; PMID:2820964
A:Accession: A29348
A:Molecule type: mRNA
A:Residues: 1725; 'V'; 1727-2409 <KRU>
A:Cross-references: GB:J02814
R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.
J. Biol. Chem. 267, 23883-23887, 1992
A:Title: Isolation of a large aggregating proteoglycan from human brain.
A:Reference number: A45131; MUID:93054750; PMID:1429726
A:Contents: brain
A:Accession: A45131
A:Molecule type: protein
A:Residues: 21-22; 'X'; 24-37 <PE2>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP:118884)
R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.
Genomics 14, 845-851, 1992
A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human
A:Reference number: I54179; MUID:93122792; PMID:1478664
A:Accession: I54179
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 251-347 <RES>
A:Cross-references: GB:S52488; NID:g263313; PIDN:AAB24878.1; PID:g263314
C:Genetics:
A:Gene: GDB:CSPG2
A:Cross-references: GDB:I27873; OMIM:118661
A:Map position: 5q12-5q14
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-2409/Product: proteoglycan 24K core protein #status predicted <MAT>
F:167-244/Domain: link protein repeat homology <LNK1>
F:265-346/Domain: link protein repeat homology <LNK2>
F:559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F:2106-2137/Domain: EGF homology <EG1>
F:2144-2175/Domain: EGF homology <EG2>
F:2182-2302/Domain: C-type lectin homology <LCH>
F:2309-2365/Domain: complement factor H repeat homology <FHD>
Query Match 79.6%; Score 43; DB 1; Length 2409;
Best Local Similarity 70.0%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
DB 2128 CVPGYSGDQC 2137
RESULT 13
T30201
Notch homolog protein - sea squirt (Halocynthia roretzi)
C:Species: Halocynthia roretzi
C:Date: 02-Sep-2000 #sequence revision 02-Sep-2000 #text change 02-Sep-2000
A:Accession: T30201
R:Hori, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.

Dev. Genes Evol. 207, 371-380, 1997

A:Title: Notch homologue from *Halocynthia roretzi* is preferentially expressed in the cerebellum
 A:Reference number: Z20775
 A:Accession: T30201
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2352 <HOR>
 A:Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA25571.1
 C:Genetics:
 A:Gene: Notch

Query Match

Best Local Similarity 77.8%; Score 42; DB 2; Length 2352;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
 |||:||||
 Db 367 CVAGYSGPRC 376

RESULT 14

T31070

notch homolog - sea urchin (*Lytechinus variegatus*)
 C:Species: *Lytechinus variegatus* (variegated urchin)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T31070
 R:Sherwood, D.R.; McClay, D.R.
 Development 124, 3363-3374, 1997
 A:Title: Identification and localization of a sea urchin Notch homologue: insights into the evolution of the Notch signaling pathway
 A:Reference number: Z20966; MUID:97454256; PMID:9310331
 A:Accession: T31070
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2531 <SHE>
 A:Cross-references: EMBL:AF000634; NID:g2570350; PID:g2570351; PIDN:AAB82088.1
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match

Best Local Similarity 77.8%; Score 42; DB 2; Length 2531;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
 |||:||||
 Db 615 CPVGTSGDNC 624

RESULT 15

T09070

probable tenascin X - mouse
 C:Species: *Mus musculus* (house mouse)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jan-2000
 C:Accession: T09070
 R:Roven, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, C.; Suck, D.
 submitted to the EMBL Data Library, October 1997
 A:Description: Sequence of the mouse major histocompatibility locus class III region.
 A:Reference number: Z16543
 A:Accession: T09070
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4006 <ROW>
 A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958
 C:Genetics:
 A:Gene: TNX
 A:Map position: 17
 A:Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1; 1501/1; 3115/1; 3208/1; 3305/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3737/3; 3822/1
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin type III
 C:Keywords: extracellular matrix
 F:422-448/Domain: EGF homology <EGF>
 F:826-906/Domain: fibronectin type III repeat homology <3FR>
 F:3789-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match

Best Local Similarity 77.8%; Score 42; DB 2; Length 4006;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Best Local Similarity 60.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CVIGXSGDRC 10
 |||:||||
 Db 532 CAVGSGDDC 541

Search completed: July 2, 2003, 07:26:26
 Job time: 14 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:02 ; Search time 10.3333 Seconds
(without alignments)
40.138 Million cell updates/sec

Title: US-09-673-785D-8
Perfect score: 54
Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	54	100.0	1217	1	EGF_MOUSE
2	49	90.7	2139	1	CRB_DROME
3	47	87.0	3381	1	PGCV_BOVIN
4	44	81.5	756	1	NRG2_MOUSE
5	44	81.5	1133	1	EGF_RAT
6	43	79.6	862	1	PGCV_MACNE
7	43	79.6	1207	1	EGF_HUMAN
8	43	79.6	3396	1	PGCV_HUMAN
9	42	77.8	1242	1	JAG1_BRARE
10	41	75.9	1064	1	FBP1_STRPU
11	41	75.9	2920	1	CLR2_MOUSE
12	40	74.1	53	1	EGF_PIG
13	40	74.1	57	1	Y402_ARCFU
14	40	74.1	2470	1	NTC2_MOUSE
15	40	74.1	2471	1	NTC2_HUMAN
16	40	74.1	2471	1	NTC2_RAT
17	39	72.2	183	1	YRF3_SHIFL
18	39	72.2	230	1	SPT_DROME
19	39	72.2	768	1	ITS8_RABIT
20	39	72.2	769	1	ITS8_HUMAN
21	39	72.2	1964	1	NTC4_MOUSE
22	39	72.2	3358	1	PGCV_RAT
23	39	72.2	3358	1	PGCV_MOUSE
24	38	70.4	117	1	YEP2_SCHPO
25	38	70.4	3106	1	LMA2_MOUSE
26	38	70.4	3672	1	LML2_CAEEL
27	38	70.4	3712	1	LMA_DROME
28	38	70.4	3718	1	LMA5_MOUSE
29	38	70.4	4544	1	LRPL_HUMAN
30	37	68.5	108	1	RODL_NEUCR
31	37	68.5	161	1	7B2_XENLA
32	37	68.5	490	1	MURE_YERPE
33	37	68.5	494	1	MURE_ECO57

34 37 68.5 494 1 MURE_ECOLI
35 37 68.5 494 1 MURE_PASMU
36 37 68.5 495 1 MURE_SALTI
37 37 68.5 495 1 MURE_SALTY
38 37 68.5 496 1 MURE_ANASP
39 37 68.5 670 1 DVLL_HUMAN
40 37 68.5 675 1 PRKS_MOUSE
41 37 68.5 799 1 TRKA_RAT
42 37 68.5 1213 1 JAG3_BRARE
43 37 68.5 2003 1 NTC4_HUMAN
44 37 68.5 2319 1 NTC3_RAT
45 37 68.5 2437 1 NTC1_BRARE

ALIGNMENTS

RESULT 1
EGF_MOUSE
ID EGF_MOUSE STANDARD; PRT; 1217 AA.
AC P01132;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].
DE EGF.
GN EGF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=66023630; PubMed=6602382;
RX Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Pong N.M., Selby M., Rutter W.J., Bell G.I.;
RT "Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.";
RL Science 221:236-240(1983).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=83219309; PubMed=6304537;
RX Gray A., Dull T.J., Ullrich A.;
RT "Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor.";
RL Nature 303:722-725(1983).
[3]
RN SEQUENCE OF 977-1029.
RP MEDLINE=73048516; PubMed=4636327;
RX Savage C.R. Jr., Inagami T., Cohen S.;
RT "The primary structure of epidermal growth factor.";
RL J. Biol. Chem. 247:7612-7621(1972).
[4]
RN DISULFIDE BONDS.
RP MEDLINE=74025498; PubMed=4750422;
RX Savage C.R. Jr., Hash J.H., Cohen S.;
RT "Epidermal growth factor. Location of disulfide bonds.";
RL J. Biol. Chem. 248:7669-7672(1973).
[5]
RN STRUCTURE BY NMR OF 977-1029.
RP MEDLINE=92118798; PubMed=1731873;
RX Montellione G.T., Wuethrich K., Burgess A.W., Nice E.C., Wagner G., Gibson K.D., Scheraga H.A.;
RT "Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints.";
RL Biochemistry 31:236-249(1992).
[6]
RN STRUCTURE BY NMR OF 977-1029.
RP MEDLINE=93075811; PubMed=1445923;
RX Konda D., Inagaki F.;
RT "Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions.";
RT

Biochemistry 31:11928-11939(1992).
 [7]
 RN STRUCTURE BY NMR OF 980-1024.
 RX MEDLINE=99180407; PubMed=10082370;
 RA Barnham K.J., Torres A.M., Alewood D., Alewood P.F., Domagala T.,
 RA Nice E.C., Norton R.S.;
 RT "Role of the 6-20 disulfide bridge in the structure and activity of
 RT epidermal growth factor.";
 RL Protein Sci. 7:1738-1749(1998).
 CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
 CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
 CC FIBROBLASTS IN CELL CULTURE.
 CC -1- SUBCELLULAR LOCATION: type I membrane protein.
 CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 1134
 CC TO 1168 DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: J00380; AAA37539.1; -
 CC EMBL: V00741; CAA24115.1; ALT_FRAME.
 CC EMBL: V00741; CAA24116.1; -
 CC FIR: A01387; EGM5MG.
 CC PDB: 1EGF; 31-JAN-94.
 CC PDB: 3EGF; 31-JAN-94.
 CC PDB: 1EPG; 31-JAN-94.
 CC PDB: 1EPH; 31-JAN-94.
 CC PDB: 1EPI; 31-JAN-94.
 CC PDB: 1EPJ; 31-JAN-94.
 CC PDB: 1EPJ; 31-JAN-94.
 CC PDB: 1A3P; 29-JUL-98.
 CC MGD: MGI:95290; Egf.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001336; EGF_1.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR000033; Ldl_receptor_rep.
 CC Pfam: PF00008; EGF; 8.
 CC Pfam: PF00058; ldl_recept_b; 7.
 CC PRINTS: PR00009; EGFTGF.
 CC SMART: SM00179; EGF_CA; 2.
 CC SMART: SM00001; EGF_like; 7.
 CC SMART: SM00135; LY; 9.
 CC PROSITE: PS00010; ASX_HYDROXYL; 3.
 CC PROSITE: PS00022; EGF_1; 1.
 CC PROSITE: PS01186; EGF_2; 6.
 CC PROSITE: PS01187; EGF_CA; 3.
 CC EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
 KW Signal; 3D-structure
 FT SIGNAL 1 28
 FT CHAIN 29 1217
 FT CHAIN 977 1029
 FT DOMAIN 29 1038
 FT TRANSMEM 1039 1058
 FT DOMAIN 1059 1217
 FT DOMAIN 327 361
 FT DOMAIN 362 402
 FT DOMAIN 403 443
 FT DOMAIN 441 483
 FT DOMAIN 747 787
 FT DOMAIN 838 876
 FT DOMAIN 877 918
 FT DOMAIN 919 959
 FT DOMAIN 978 1019
 FT DISULFID 366 377
 FT DISULFID 373 386
 FT DISULFID 388 401
 FT DISULFID 407 418

FT DISULFID 414 427
 FT DISULFID 429 442
 FT DISULFID 445 457
 FT DISULFID 453 467
 FT DISULFID 469 482
 FT DISULFID 751 762
 FT DISULFID 758 771
 FT DISULFID 773 786
 FT DISULFID 842 853
 FT DISULFID 864 875
 FT DISULFID 881 895
 FT DISULFID 888 904
 FT DISULFID 906 917
 FT DISULFID 923 936
 FT DISULFID 930 945
 FT DISULFID 947 958
 FT DISULFID 982 996
 FT DISULFID 990 1007
 FT DISULFID 1009 1018
 FT DOMAIN 1024 1029
 FT CARBOHYD 111 111
 FT CARBOHYD 410 410
 FT CARBOHYD 810 810
 FT CARBOHYD 944 944
 FT CONFLICT 790 790
 FT CONFLICT 1048 1048
 FT STRAND 995 997
 FT STRAND 1006 1008
 FT STRAND 1010 1010
 FT TURN 1011 1012
 FT STRAND 1013 1014
 FT STRAND 1020 1021
 SQ SEQUENCE 1217 AA; 133144 MW; A9C7F3D512F82873 CRC64;
 Query Match 100.0%; Score 54; DB 1; Length 1217;
 Best Local Similarity 90.0%; Pred. No. 0.05;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDRC 10
 Db 1009 CVIGXSGDRC 1018
 |||||
 RESULT 2
 ID CRB_DROME STANDARD; PRT; 2139 AA.
 AC P10040;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Crumbs protein precursor (95F).
 GN CRB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Oregon-R; TISSUE-Embryo;
 RX MEDLINE=90263104; PubMed=2344615;
 RA Tepass U., Theres C., Knust E.;
 RT "Crumbs encodes an EGF-like protein expressed on apical membranes of
 RT Drosophila epithelial cells and required for organization of
 RT epithelia.";
 RL Cell 61:787-799(1990).
 RN [2]
 RP SEQUENCE OF 1663-1955 FROM N.A.
 RC TISSUE-Embryo;
 RX MEDLINE=87218537; PubMed=3107986;
 RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,

NOT REQUIRED FOR FULL BIOLOGICAL ACTIVITY
 IN VIVO.
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 D -> Y (IN REF. 2).
 A -> S (IN REF. 2).

RA Vaessin H., Campos-Ortega J.A.;
 RT "EGF homologous sequences encoded in the genome of Drosophila
 RL melanogaster, and their relation to neurogenic genes."; EMBO J. 6:761-766(1987).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
 CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
 CC POLARITY. IT MAY ACT AS A SIGNAL.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
 CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M33753; AAA28428.1; ALT_SEQ.
 DR EMBL; X05144; CAA28793.1; -.
 DR PIR; B26637; B26637.
 DR PIR; A35672; A35672.
 DR HSP; P00740; IEDM.
 DR FlyBase; FBgn0000368; crb.
 DR InterPro; IPR00152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001861; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 26.
 DR Pfam; PF00054; laminin_G; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00001; EGF_like; 16.
 DR SMART; SM00282; LamG; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 15.
 DR PROSITE; PS00022; EGF_1; 26.
 DR PROSITE; PS01186; EGF_2; 17.
 DR PROSITE; PS01187; EGF_CA; 12.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 90
 FT CHAIN 91 2139 CRUMBS PROTEIN.
 FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2085 2111 POTENTIAL.
 FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 267 303 EGF-LIKE 1.
 FT DOMAIN 306 343 EGF-LIKE 2.
 FT DOMAIN 348 386 EGF-LIKE 3.
 FT DOMAIN 388 425 EGF-LIKE 4.
 FT DOMAIN 427 463 EGF-LIKE 5.
 FT DOMAIN 464 500 EGF-LIKE 6.
 FT DOMAIN 501 532 EGF-LIKE 7.
 FT DOMAIN 545 581 EGF-LIKE 8.
 FT DOMAIN 582 611 EGF-LIKE 9.
 FT DOMAIN 609 646 EGF-LIKE 10.
 FT DOMAIN 648 685 EGF-LIKE 11.
 FT DOMAIN 687 723 EGF-LIKE 12.
 FT DOMAIN 725 761 EGF-LIKE 13.
 FT DOMAIN 763 800 EGF-LIKE 14.
 FT DOMAIN 802 838 EGF-LIKE 15.
 FT DOMAIN 840 902 EGF-LIKE 16.
 FT DOMAIN 904 948 EGF-LIKE 17.
 FT DOMAIN 942 978 EGF-LIKE 18.
 FT DOMAIN 980 1021 EGF-LIKE 19.
 FT DOMAIN 1023 1205 LAMININ G-LIKE 1.
 FT DOMAIN 1207 1243 EGF-LIKE 20.
 FT DOMAIN 1250 1480 LAMININ G-LIKE 2.
 FT DOMAIN 1481 1517 EGF-LIKE 21.

FT DOMAIN	1558	1758	LAMININ G-LIKE 3.
FT DOMAIN	1759	1795	EGF-LIKE 22.
FT DOMAIN	1797	1833	EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	1835	1871	EGF-LIKE 24. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	1874	1915	EGF-LIKE 25.
FT DOMAIN	1915	1951	EGF-LIKE 26.
FT DOMAIN	1953	1989	EGF-LIKE 27. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	1991	2029	EGF-LIKE 28. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	2030	2070	EGF-LIKE 29.
FT DISULFID	271	282	BY SIMILARITY.
FT DISULFID	276	291	BY SIMILARITY.
FT DISULFID	293	302	BY SIMILARITY.
FT DISULFID	310	321	BY SIMILARITY.
FT DISULFID	333	341	BY SIMILARITY.
FT DISULFID	342	351	BY SIMILARITY.
FT DISULFID	352	363	BY SIMILARITY.
FT DISULFID	357	374	BY SIMILARITY.
FT DISULFID	376	385	BY SIMILARITY.
FT DISULFID	392	403	BY SIMILARITY.
FT DISULFID	397	412	BY SIMILARITY.
FT DISULFID	414	424	BY SIMILARITY.
FT DISULFID	431	442	BY SIMILARITY.
FT DISULFID	436	451	BY SIMILARITY.
FT DISULFID	453	462	BY SIMILARITY.
FT DISULFID	468	479	BY SIMILARITY.
FT DISULFID	473	488	BY SIMILARITY.
FT DISULFID	490	499	BY SIMILARITY.
FT DISULFID	505	515	BY SIMILARITY.
FT DISULFID	509	520	BY SIMILARITY.
FT DISULFID	522	531	BY SIMILARITY.
FT DISULFID	549	562	BY SIMILARITY.
FT DISULFID	556	569	BY SIMILARITY.
FT DISULFID	571	580	BY SIMILARITY.
FT DISULFID	586	597	BY SIMILARITY.
FT DISULFID	591	602	BY SIMILARITY.
FT DISULFID	604	610	BY SIMILARITY.
FT DISULFID	613	624	BY SIMILARITY.
FT DISULFID	618	634	BY SIMILARITY.
FT DISULFID	636	645	BY SIMILARITY.
FT DISULFID	652	664	BY SIMILARITY.
FT DISULFID	659	673	BY SIMILARITY.
FT DISULFID	675	684	BY SIMILARITY.
FT DISULFID	691	702	BY SIMILARITY.
FT DISULFID	696	711	BY SIMILARITY.
FT DISULFID	713	722	BY SIMILARITY.
FT DISULFID	729	740	BY SIMILARITY.
FT DISULFID	734	749	BY SIMILARITY.
FT DISULFID	751	760	BY SIMILARITY.
FT DISULFID	767	778	BY SIMILARITY.
FT DISULFID	772	787	BY SIMILARITY.
FT DISULFID	789	799	BY SIMILARITY.
FT DISULFID	806	817	BY SIMILARITY.
FT DISULFID	811	826	BY SIMILARITY.
FT DISULFID	828	837	BY SIMILARITY.
FT DISULFID	844	855	BY SIMILARITY.
FT DISULFID	849	890	BY SIMILARITY.
FT DISULFID	892	901	BY SIMILARITY.
FT DISULFID	908	919	BY SIMILARITY.
FT DISULFID	913	928	BY SIMILARITY.
FT DISULFID	930	939	BY SIMILARITY.
FT DISULFID	946	957	BY SIMILARITY.
FT DISULFID	952	966	BY SIMILARITY.
FT DISULFID	968	977	BY SIMILARITY.
FT DISULFID	984	995	BY SIMILARITY.
FT DISULFID	989	1009	BY SIMILARITY.
FT DISULFID	1011	1020	BY SIMILARITY.
FT DISULFID	1211	1222	BY SIMILARITY.
FT DISULFID	1216	1231	BY SIMILARITY.
FT DISULFID	1233	1242	BY SIMILARITY.
FT DISULFID	1485	1496	BY SIMILARITY.
FT DISULFID	1490	1505	BY SIMILARITY.
FT DISULFID	1507	1516	BY SIMILARITY.
FT DISULFID	1763	1774	BY SIMILARITY.

FT	DOMAIN	1337	3074	GAG-BETA.	DE
FT	DOMAIN	3074	3110	EGF-LIKE 1.	GN
FT	DOMAIN	3112	3148	EGF-LIKE 2, CALCIUM-BINDING.	OS
FT	DOMAIN	3161	3275	C-TYPE LECTIN.	OC
FT	DOMAIN	3280	3338	SUSHI.	OC
FT	DISULFID	44	131	BY SIMILARITY.	OX
FT	DISULFID	173	244	BY SIMILARITY.	RN
FT	DISULFID	197	218	BY SIMILARITY.	RP
FT	DISULFID	271	346	BY SIMILARITY.	RC
FT	DISULFID	295	316	BY SIMILARITY.	RX
FT	DISULFID	3078	3089	BY SIMILARITY.	RA
FT	DISULFID	3083	3098	BY SIMILARITY.	RA
FT	DISULFID	3100	3109	BY SIMILARITY.	RT
FT	DISULFID	3121	3127	BY SIMILARITY.	RT
FT	DISULFID	3126	3136	BY SIMILARITY.	RL
FT	DISULFID	3138	3147	BY SIMILARITY.	RN
FT	DISULFID	3154	3165	BY SIMILARITY.	RP
FT	DISULFID	3182	3274	BY SIMILARITY.	RC
FT	DISULFID	3250	3266	BY SIMILARITY.	RX
FT	DISULFID	3281	3324	BY SIMILARITY.	RA
FT	DISULFID	3310	3337	BY SIMILARITY.	RA
FT	CARBOHYD	57	57	N-LINKED (GLCNAC. . .)	RA
FT	CARBOHYD	331	331	N-LINKED (GLCNAC. . .)	RT
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .)	RT
FT	CARBOHYD	817	817	N-LINKED (GLCNAC. . .)	RL
FT	CARBOHYD	965	965	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1017	1017	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1333	1333	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1393	1393	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1437	1437	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1463	1463	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1653	1653	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	1974	1974	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2045	2045	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2074	2074	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2103	2103	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2263	2263	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2356	2356	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2623	2623	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2641	2641	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	2919	2919	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	3354	3354	N-LINKED (GLCNAC. . .)	CC
FT	CARBOHYD	3364	3364	N-LINKED (GLCNAC. . .)	CC
FT	VARSPLIC	349	349	P -> R (IN ISOFORM V1 AND V3).	CC
FT	VARSPLIC	350	1336	MISSING (IN ISOFORM V1).	CC
FT	VARSPLIC	1337	3074	MISSING (IN ISOFORM V2).	CC
FT	CONFLICT	25	25	MISSING (IN ISOFORM V3).	CC
FT	CONFLICT	51	51	MISSING (IN REF. 2).	CC
FT	CONFLICT	89	89	N -> D (IN REF. 2).	CC
FT	CONFLICT	96	96	O -> D (IN REF. 2).	CC
FT	CONFLICT	346	346	C -> R (IN REF. 2).	DR
SQ	SEQUENCE	3381 AA;	369984 MW;	F09716FA7778D459 CRC64;	DR
Query Match				87.0%; Score 47; DB 1; Length 3381;	DR
Best Local Similarity				80.0%; Pred. No. 2.6;	DR
Matches				8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	DR
QY	1	CVIGXSGDRC	10		DR
Db	3100	CVPGYSGDRC	3109		DR
RESULT 4					DR
ID	NRG2_MOUSE	STANDARD;	PRT;	756 AA.	DR
AC	P56974;				DR
DT	16-OCT-2001 (Rel. 40, Created)				DR
DT	16-OCT-2001 (Rel. 40, Last sequence update)				DR
DT	16-OCT-2001 (Rel. 40, Last annotation update)				DR
DE	Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)]				DR

(Divergent of neuregulin 1) (DON-1)).

NRG2.

Mus musculus (Mouse).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

[1]

SEQUENCE FROM N.A. (ISOFORMS NRG2-5; NRG2-10 AND NRG2-16A).

STRAIN=C57BL/6; TISSUE=Brain;

MEDLINE=973111398; PubMed=9168115;

Carraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N., Gassmann M., Lai C.;

"Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases.";

Nature 387:512-516(1997).

[2]

SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).

TISSUE=Choroid plexus;

MEDLINE=97342638; PubMed=9199335;

Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J., Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J., Gearing D.P.;

"Characterization of a neuregulin-related gene, Don-1, that is highly expressed in restricted regions of the cerebellum and hippocampus.";

Mol. Cell. Biol. 17:4007-4014(1997).

-1- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS, RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE HETERODIMERIZATION WITH THE EGF RECEPTOR.

-1- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).

-1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS: DON-1M, DON-1S/NRG2-5, NRG2-10 AND NRG2-16A (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WITH LOWER LEVELS IN THE LUNG. IN THE CEREBELLUM, FOUND IN GRANULE AND PURKINJE CELLS.

-1- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN DIMERIZATION (BY SIMILARITY).

-1- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE DOMAIN (BY SIMILARITY).

-1- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR FORM (BY SIMILARITY).

-1- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY SIMILARITY).

-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.

-1- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.

HSSP: Q12784; IHR.

MGD; MGI:1098246; Nrg2.

InterPro: IPR000561; EGF-like.

InterPro: IPR003006; Ig_MHC.

InterPro: IPR003598; Ig_c2.

InterPro: IPR002154; Neuregulin.

Pfam: PF000008; EGF; 1.

Pfam: PF000047; Ig; 1.

Pfam: PF02158; Neuregulin; 1.

SMART: SM00181; EGF; 1.

SMART: SM00408; IGC2; 1.

PROSITE: PS00022; EGF_1; 1.

PROSITE: PS01186; EGF_2; 1.

Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein; Transmembrane; Multigene family; Alternative splicing.

PROPEP 1 19

CHAIN 20 756

CHAIN 20 314

DOMAIN 20 315

TRANSMEM 316 336

PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.

NEUREGULIN-2.

EXTRACELLULAR (POTENTIAL).

INTERNAL SIGNAL SEQUENCE (POTENTIAL).

```

FT DOMAIN 337 756 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 226 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 238 248 SER/THR-RICH.
FT DOMAIN 290 290 EGF-LIKE.
FT DOMAIN 627 633 POLY-PRO.
FT DISULFID 155 219 BY SIMILARITY.
FT DISULFID 253 267 BY SIMILARITY.
FT DISULFID 261 278 BY SIMILARITY.
FT DISULFID 280 289 BY SIMILARITY.
FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 C -> G (IN ISOFORM NR2-10).
FT VARSPLIC 280 280 MISSING (IN ISOFORM NR2-10).
FT VARSPLIC 281 756 VGYTGDCQQQFAMVNFSKHLGFEKEAEELYOKRVLTITGI
FT VARSPLIC 282 330 CVALIIVG -> NGFFGQRCLEKPLRLYMDPKQSVLWMDT
FT VARSPLIC 331 756 PGTGVSSQWSTSTSLDNLN (IN ISOFORM DON-1S).
FT VARSPLIC 282 307 VGYTGDCQQQFAMVNFSKHLGFEKE -> NGFFGQRCLEK
SQ SEQUENCE 756 AA; 82213 MW; 51D85DC918B678E CRC64;

Query Match 81.5%; Score 44; DB 1; Length 756;
Best Local Similarity 60.0%; Pred. No. 2.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
| : : : : :
Db 280 CPVGYTGDC 289

RESULT 5
EGF_RAT
ID EGF_RAT STANDARD; PRT; 1133 AA.
AC P07522; Q63183;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].
DE EGF.
GN EGF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-Kidney;
RC MEDLINE=92398779; PubMed=1524680;
RA Price P.M., Sagai S.J., Safirstein R.;
RT "Cloning and sequencing of the rat preproepidermal growth factor
cDNA: comparison with mouse and human sequences.";
RL DNA Cell Biol. 11:481-487(1992).
[2]
RN REVISIONS.
RP TISSUE-Kidney;
RC Price P.M.;
RA Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 974-1021.
RX MEDLINE=86081810; PubMed=3000782;
RA Simpson R.J., Smith J.A., Moritz R.L., O'Hare M.J., Rudland P.S.,
RA Morrison J.R., Lloyd C.J., Grego B., Burgess A.W., Nice E.C.;
RT "Rat epidermal growth factor: complete amino acid sequence. Homology
with the corresponding murine and human proteins; isolation of a form
truncated at both ends with full in vitro biological activity.";
RL Eur. J. Biochem. 153:629-637(1985).
[4]
RN SEQUENCE OF 994-1108 FROM N.A.
RP STRAIN-Sprague-Dawley; TISSUE-Kidney;
RX MEDLINE=89016634; PubMed=3262867;
RA Dorow D.S., Simpson R.J.;

```

```

RT "Cloning and sequence analysis of a cDNA for rat epidermal growth
factor.";
RL Nucleic Acids Res. 16:9338-9338(1988).
CC FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC SUBCELLULAR LOCATION: Type I membrane protein.
CC SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U04842; AAB60436.1; -.
DR EMBL; X12748; CAA31241.1; -.
DR PIR; A25425; EGR1.
DR HSSP; P01132; IEGF.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001336; EGF-1.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 7.
DR Pfam; PF00058; ldl_recept_b; 7.
DR PRINTS; PR00009; EGF1.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF-like; 6.
DR SMART; SM00135; LY; 8.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 3.
KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
KW Signal.
FT CHAIN 1 21 POTENTIAL.
FT CHAIN 22 1133 PRO-EPIDERMAL GROWTH FACTOR.
FT DOMAIN 974 1026 EPIDERMAL GROWTH FACTOR.
FT DOMAIN 22 1035 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1036 1057 POTENTIAL.
FT DOMAIN 1058 1133 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 322 356 EGF-LIKE 1 (INCOMPLETE).
FT DOMAIN 357 397 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 398 438 EGF-LIKE 3.
FT DOMAIN 436 478 EGF-LIKE 4.
FT DOMAIN 743 783 EGF-LIKE 5.
FT DOMAIN 835 873 EGF-LIKE 6.
FT DOMAIN 874 915 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 916 956 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 975 1016 EGF-LIKE 9.
FT DISULFID 361 372 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 383 396 BY SIMILARITY.
FT DISULFID 402 413 BY SIMILARITY.
FT DISULFID 409 422 BY SIMILARITY.
FT DISULFID 424 437 BY SIMILARITY.
FT DISULFID 440 452 BY SIMILARITY.
FT DISULFID 448 462 BY SIMILARITY.
FT DISULFID 464 477 BY SIMILARITY.
FT DISULFID 747 758 BY SIMILARITY.
FT DISULFID 754 767 BY SIMILARITY.
FT DISULFID 769 782 BY SIMILARITY.
FT DISULFID 839 850 BY SIMILARITY.
FT DISULFID 844 859 BY SIMILARITY.
FT DISULFID 861 872 BY SIMILARITY.
FT DISULFID 878 892 BY SIMILARITY.
FT DISULFID 885 901 BY SIMILARITY.
FT DISULFID 903 914 BY SIMILARITY.
FT DISULFID 920 933 BY SIMILARITY.
FT DISULFID 927 942 BY SIMILARITY.

```

```

FT DISULFID 944 955 BY SIMILARITY.
FT DISULFID 979 993
FT DISULFID 987 1004
FT DISULFID 1006 1015
FT CARBOHYD 105 108
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 955 955 C -> V.
FT VARIANT 1024 1025 KL -> S (IN REF. 4).
FT CONFLICT 1108 1108 Q -> S (IN REF. 4).
SQ SEQUENCE 1133 AA; 124125 MW; C224A302E9578031 CRC64;

Query Match 81.5%; Score 44; DB 1; Length 1133;
Best Local Similarity 70.0%; Pred. No. 3.1; Mismatches 1; Indels 0; Gaps 0;
Matches 7; Conservative 2;

QY 1 CVIGXSGDRC 10
Db 1006 CVIGIGERC 1015
||||:|:|

RESULT 6
PGCV_MACNE STANDARD; PRT; 862 AA.
AC Q28658; Q28859; Q28860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Versican core protein (large fibroblast proteoglycan) (Chondroitin
DE sulfate proteoglycan core protein 2) (Fragments).
GN CSPG2.
OS Macaca nemestrina (Pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9545;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Aortic smooth muscle;
RA MEDLINE=95005762; PubMed=7921538;
RX Yao L.Y., Moody C., Schoenherr E., Wight T.N., Sandell L.J.;
RT "Identification of the proteoglycan versican in aorta and smooth
RT muscle cells by DNA sequence analysis, in situ hybridization and
RT immunohistochemistry.";
RL Matrix Biol. 14:213-225(1994).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronan.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; S72412; AAA65593.2; -
CC EMBL; S72413; AAA65594.2; -
CC EMBL; S72414; AAA65595.2; -
CC HSP; P01132; IEGF.

```

```

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000538; Link.
DR PRINTS: PR00010; EGFBL00D.
DR ProDom: PD000918; Link; 2.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00445; LINK; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01241; LINK; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; PARTIAL.
DR PROSITE: PS0041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; proteoglycan; Lectin; Extracellular matrix; Repeat;
KW EGF-like domain; Calcium.
FT NON_TER 1 1
FT DOMAIN <1 37 LINK 1.
FT DOMAIN 58 139 LINK 2.
FT DOMAIN 141 >233 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT DOMAIN) .
FT NON_CONS 233 234
FT DOMAIN <234 >525 GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN,
FT SIMILAR TO CHONDROITIN SULFATE ATTACHMENT
FT SITE IN COLLAGEN TYPE IX (BY SIMILARITY).
FT EGF-LIKE 1.
FT EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT C-TYPE LECTIN.
FT BY SIMILARITY.
FT DISULFID 63 138 BY SIMILARITY.
FT DISULFID 87 108 BY SIMILARITY.
FT DISULFID 722 733 BY SIMILARITY.
FT DISULFID 727 742 BY SIMILARITY.
FT DISULFID 744 753 BY SIMILARITY.
FT DISULFID 760 771 BY SIMILARITY.
FT DISULFID 765 780 BY SIMILARITY.
FT DISULFID 782 791 BY SIMILARITY.
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 563 563 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 862 862
SQ SEQUENCE 862 AA; 95583 MW; A5D5F6153A74BB39 CRC64;

Query Match 79.6%; Score 43; DB 1; Length 862;
Best Local Similarity 70.0%; Pred. No. 3.6;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 744 CVPYSGDQC 753
||||:|:|

RESULT 7
EGF_HUMAN STANDARD; PRT; 1207 AA.
ID EGF_HUMAN
AC P01133;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal
DE growth factor (Urogastrone)].
GN EGF.

```

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=87066721; PubMed=3491360;
 RA Bell G.I., Fong N.M., Stempien M.M., Wormsted M.A., Caput D.,
 RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;
 RA "Human epidermal growth factor precursor: cDNA sequence, expression
 RT in vitro and gene organization.";
 RL Nucleic Acids Res. 14:8427-8446(1986).
 RN [2]
 RP SEQUENCE OF 971-1023.
 RX MEDLINE=77117897; PubMed=300079;
 RA Gregory H., Preston B.M.;
 RA "The primary structure of human urogastrone.";
 RT Int. J. Pept. Protein Res. 9:107-118(1977).
 RN [3]
 RP SEQUENCE OF 971-1023.
 RX MEDLINE=89391964; PubMed=2789514;
 RA Furuya M., Akashi S., Hirayama K.;
 RA "The primary structure of human EGF produced by genetic engineering,
 RT studied by high-performance tandem mass spectrometry.";
 RL Biochem. Biophys. Res. Commun. 163:1100-1106(1989).
 RN [4]
 RP STRUCTURE BY NMR OF EGF.
 RX MEDLINE=92395667; PubMed=1522591;
 RA Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;
 RA "Human epidermal growth factor. High resolution solution structure
 RT and comparison with human transforming growth factor alpha.";
 RL J. Mol. Biol. 227:271-282(1992).
 CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
 CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
 CC FIBROBLASTS IN CELL CULTURE.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X04571; CAA28240.1; -
 CC PIR: A25531; EGHU
 CC HSP; P01132; IEGF.
 CC Genew; HGNC:3229; EGF.
 CC MIM; 131530; -
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR001336; EGF_1.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR000033; Ild_receptor_rep.
 CC Pfam; PF00008; EGF; 9.
 CC Pfam; PF00058; Ild_recept_b; 7.
 CC PRINTS; PR00009; EGF_TGF.
 CC SMART; SM00179; EGF_CA; 2.
 CC SMART; SM00001; EGF-like; 7.
 CC SMART; SM00135; LY; 8.
 CC PROSITE; PS00010; ASX_HYDROXYL; 3.
 CC PROSITE; PS00022; EGF_1; 1.
 CC PROSITE; PS01186; EGF_2; 7.
 CC PROSITE; PS01187; EGF_CA; 3.
 CC EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
 KW Signal; Polymorphism.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 1207 PRO-EPIDERMAL GROWTH FACTOR.
 FT CHAIN 971 1023 EPIDERMAL GROWTH FACTOR.
 FT DOMAIN 23 1032 EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	1033	1053	
FT	DOMAIN	1054	1207	POTENTIAL.
FT	DOMAIN	314	355	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	356	396	EGF-LIKE 1. CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	397	437	EGF-LIKE 2.
FT	DOMAIN	435	477	EGF-LIKE 3.
FT	DOMAIN	741	781	EGF-LIKE 4.
FT	DOMAIN	831	869	EGF-LIKE 5.
FT	DOMAIN	870	911	EGF-LIKE 6.
FT	DOMAIN	912	952	EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	972	1013	EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	318	330	BY SIMILARITY.
FT	DISULFID	325	339	BY SIMILARITY.
FT	DISULFID	341	354	BY SIMILARITY.
FT	DISULFID	360	371	BY SIMILARITY.
FT	DISULFID	367	380	BY SIMILARITY.
FT	DISULFID	382	395	BY SIMILARITY.
FT	DISULFID	401	412	BY SIMILARITY.
FT	DISULFID	408	421	BY SIMILARITY.
FT	DISULFID	423	436	BY SIMILARITY.
FT	DISULFID	439	451	BY SIMILARITY.
FT	DISULFID	447	461	BY SIMILARITY.
FT	DISULFID	463	476	BY SIMILARITY.
FT	DISULFID	745	756	BY SIMILARITY.
FT	DISULFID	767	780	BY SIMILARITY.
FT	DISULFID	835	846	BY SIMILARITY.
FT	DISULFID	840	855	BY SIMILARITY.
FT	DISULFID	857	868	BY SIMILARITY.
FT	DISULFID	874	888	BY SIMILARITY.
FT	DISULFID	889	897	BY SIMILARITY.
FT	DISULFID	916	929	BY SIMILARITY.
FT	DISULFID	923	938	BY SIMILARITY.
FT	DISULFID	940	951	BY SIMILARITY.
FT	DISULFID	976	990	
FT	DISULFID	984	1001	
FT	DISULFID	1003	1012	
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	104	104	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	117	117	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	148	148	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	324	324	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	404	404	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	596	596	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	815	815	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	926	926	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	708	708	I -> M.
FT	SEQUENCE	1207 AA; 133945 MW; D627DC828EF782E9 CRC64;		/FTid=VAR_002275.
Query Match		79.68;		
Best Local Similarity		60.08;		
Matches	6;	Conservative	3;	Mismatches 1; Indels 0; Gaps 0;

RESULT 8

PGCV_HUMAN
 ID PGCV_HUMAN STANDARD; PRT; 3396 AA.
 AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
 DE hyaluronate-binding protein) (GHP).
 GN CSFG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RP [1]
RX SEQUENCE FROM N.A. (ISOFORM V0).
RA MEDLINE=95105188; PubMed=7528742;
RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT "Characterization of the complete genomic structure of the human
RT versican gene and functional analysis of its promoter";
RL J. Biol. Chem. 269:32999-33008(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V1).
RX TISSUE=Placenta; PubMed=2583089;
RX MEDLINE=90059882; PubMed=2583089;
RA Zimmermann D.R., Ruoslahti E.;
RT "Multiple domains of the large fibroblast proteoglycan, versican";
RL EMBO J. 8:2975-2981(1989).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM V2).
RX TISSUE=Glial tumor;
RX MEDLINE=95105187; PubMed=7806529;
RA Dours-Zimmermann M.T., Zimmermann D.R.;
RT "A novel glycosaminoglycan attachment domain identified in two
RT alternative splice variants of human versican";
RL J. Biol. Chem. 269:32992-32998(1994).
RN [4]
RP SEQUENCE OF 2711-3396 FROM N.A.
RX TISSUE=Lung fibroblast;
RX MEDLINE=88007514; PubMed=2820964;
RA Krusius T., Gehlsen K.R., Ruoslahti E.;
RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
RT lectin-like and growth factor-like sequences";
RL J. Biol. Chem. 262:13120-13125(1987).
RN [5]
RP SEQUENCE OF 251-347 FROM N.A.
RX MEDLINE=93122792; PubMed=1478664;
RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
RA McPherson J.D.;
RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RT human chromosome 5 (5q12-5q14)";
RL Genomics 14:845-851(1992).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM V3).
RX TISSUE=Brain;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues";
RL J. Biol. Chem. 270:3914-3918(1995).
RN [7]
RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RX TISSUE=Aortic smooth muscle;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [8]
RP PARTIAL SEQUENCE.
RX TISSUE=Brain;
RX MEDLINE=89174663; PubMed=2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial
RT hyaluronate-binding protein";
RL J. Biol. Chem. 264:5981-5987(1989).
RN [9]
RP TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE=96213482; PubMed=8627343;
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT "Differential expression of versican isoforms in brain tumors";
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
CC -!- FUNCTION: May play a role in intercellular signaling and in
connecting cells with the extracellular matrix. May take part in

CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 5 isoforms: V0 (shown here), V1,
CC V2, V3 and Vint; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
CC in normal brain, gliomas, medulloblastomas, schwannomas,
CC neurofibromas, and meningiomas; V2 is restricted to normal brain
CC and gliomas; V3 is found in all these tissues except
CC medulloblastomas.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U16306; AAA65018.1; -
CC EMBL: X15998; CAA34128.1; -
CC EMBL: S52488; AAB24878.1; -
CC EMBL: U28555; AAA67565.1; -
CC EMBL: J32039; BAA06801.1; -
CC EMBL: J02814; AAA36437.1; -
CC EMBL: AF084545; AAD48545.1; -
CC PIR: S06014; S06014.
CC PIR: A29348; A29348.
CC PIR: A30358; A30358.
CC HSP: P01132; IEGF.
CC Genew: HGNC:2464; CSPG2.
CC MIM: 118661; -
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF_2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR001438; EGF_II.
CC InterPro: IPR003599; Ig.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR001304; Lectin_C.
CC InterPro: IPR000538; Link.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: PF00008; EGF; 2.
CC Pfam: PF00047; Ig; 1.
CC Pfam: PF00059; Lectin_C; 1.
CC Pfam: PF00084; sushi; 1.
CC Pfam: PF00193; Xlink; 2.
CC PRINTS: PR00010; EGF_BLOOD.
CC PRODOM: PD000918; Link; 2.
CC SMART: SM00032; CCP; 1.
CC SMART: SM00034; CLECT; 1.
CC SMART: SM00179; EGF_CA; 1.
CC SMART: SM00001; EGF_like; 1.
CC SMART: SM00409; IG; 1.
CC SMART: SM00445; LINK; 2.
CC PROSITE: PS00010; ASX_HYDROXYL; 1.
CC PROSITE: PS00022; EGF_1; 2.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS01187; EGF_CA; 1.
CC PROSITE: PS01241; LINK; 2.
CC PROSITE: PS00615; C-TYPE LECTIN; 1.
CC PROSITE: PS00041; C-TYPE LECTIN; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20
FT POTENTIAL.


```
FT DISULFID 476 485 BY SIMILARITY.
FT DISULFID 492 503 BY SIMILARITY.
FT DISULFID 497 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 751 762 BY SIMILARITY.
FT DISULFID 756 771 BY SIMILARITY.
FT DISULFID 773 782 BY SIMILARITY.
FT DISULFID 789 800 BY SIMILARITY.
FT DISULFID 794 809 BY SIMILARITY.
FT DISULFID 811 820 BY SIMILARITY.
FT DISULFID 827 838 BY SIMILARITY.
FT DISULFID 832 847 BY SIMILARITY.
FT DISULFID 849 858 BY SIMILARITY.
FT CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 385 385 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 560 560 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1242 AA; 135343 MW; 59557A4B7F052860 CRC64;

Query Match 77.8%; Score 42; DB 1; Length 1242;
Best Local Similarity 60.0%; Pred. No. 8;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
   1 1::1111
DB 476 CAAGFAGDRC 485

RESULT 10
FBP1_STRPU STANDARD; PRT; 1064 AA.
AC P10079;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibropellin I precursor (Epidermal growth factor-related protein 1)
DE (UEGF-1).
GN EGF1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_Taxid=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90112459; PubMed=2514273;
RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;
RT "Structural analysis of the uegf gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats.";
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE=87319677; PubMed=3498216;
RA Hursh D.A., Andrews M.E., Raff R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor.";
```

```
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE=89196806; PubMed=2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin.";
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91285254; PubMed=2060714;
RA Bisgrove B.W., Andrews M.E., Raff R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo.";
RL Dev. Biol. 146:89-99(1991).
CC -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF
CC REPEATS.
CC -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08692; AA62164.1; -
DR EMBL; L08692; AA62163.1; -
DR EMBL; X17530; CAA35571.1; -
DR EMBL; M17421; AAA30050.1; -
DR EMBL; X17533; CAA35573.1; -
DR PIR; A29316; A29316.
DR HSP; P01132; 1EGF.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000088; Avidin.
DR InterPro; IPR000859; CUB.domain.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00008; EGF; 21.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF01382; Avidin; 1.
DR PRINTS; PR00010; EGFBLAMIN.
DR PRINTS; PR00011; EGFBLAMIN.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00179; EGF_CA; 18.
DR SMART; SM00001; EGF-like; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 19.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS01187; EGF_CA; 18.
KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
```

KW Glycoprotein; Calcium-binding.
FT SIGNAL 1 19
FT CHAIN 20 1064 POTENTIAL.
FT DOMAIN 20 55 FIBROPELIN I.
FT DOMAIN 62 175 EGF-LIKE 1.
FT DOMAIN 176 212 CUB.
FT DOMAIN 212 250 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 860 896 EGF-LIKE 20.
FT DOMAIN 898 934 AVIDIN-LIKE.
FT DOMAIN 936 1064 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 23 34 BY SIMILARITY.
FT DISULFID 28 43 BY SIMILARITY.
FT DISULFID 45 54 BY SIMILARITY.
FT DISULFID 180 191 BY SIMILARITY.
FT DISULFID 185 200 BY SIMILARITY.
FT DISULFID 202 211 BY SIMILARITY.
FT DISULFID 218 229 BY SIMILARITY.
FT DISULFID 223 238 BY SIMILARITY.
FT DISULFID 240 249 BY SIMILARITY.
FT DISULFID 256 267 BY SIMILARITY.
FT DISULFID 261 276 BY SIMILARITY.
FT DISULFID 278 305 BY SIMILARITY.
FT DISULFID 294 307 BY SIMILARITY.
FT DISULFID 319 325 BY SIMILARITY.
FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 337 352 BY SIMILARITY.
FT DISULFID 354 363 BY SIMILARITY.
FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 603 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 750 761 BY SIMILARITY.
FT DISULFID 755 770 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 793 808 BY SIMILARITY.
FT DISULFID 810 819 BY SIMILARITY.
FT DISULFID 826 837 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 848 857 BY SIMILARITY.
FT DISULFID 864 875 BY SIMILARITY.
FT DISULFID 869 884 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 902 913 BY SIMILARITY.
FT DISULFID 907 922 BY SIMILARITY.
FT DISULFID 924 933 BY SIMILARITY.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 477 780 MISSING (IN ISOFORM IB).
FT CONFLICT 279 279 L -> S (IN REF. 2).
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;
Query Match 75.9%; Score 41; DB 1; Length 1064;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 CVIGXSGDRC 10
DB 316 CPLGFGDNC 325
RESULT 11
CLR2_MOUSE STANDARD; PRT; 2920 AA.
ID CLR2_MOUSE Q9Z2R4; Q99K26;
AC Q9ROM0; Q9Z2R4; Q99K26;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)
DE (mFml1).
GN CELSR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99418630; PubMed=10490098;
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
polarity under the control of frizzled.";
RL Cell 98:585-595(1999).
RN [2]
RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.
RX PubMed=10790539;
RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;
RT "Chromosomal localization of Celser2 and Celser3 in the mouse: Celser3 is
a candidate for the tippy (tip) lethal mutant on chromosome 9.";
RL Mamm. Genome 11:392-394(2000).
RN [3]
RP SEQUENCE OF 2014-2920 FROM N.A.
RC TISSUE-Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP DEVELOPMENTAL STAGE.
RX PubMed=11850187;
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celser (Flamingo) genes in the
mouse.";
RL Mech. Dev. 112:157-160(2002).


```
FT CARBOHYD 1502 1502 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1566 1566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1742 1742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1828 1828 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1901 1901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2025 2025 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2062 2062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2324 2324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2346 2346 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2199 2199 V -> L (IN REF. 3).
FT CONFLICT 2283 2283 A -> V (IN REF. 3).
FT CONFLICT 2535 2535 S -> R (IN REF. 1).
FT CONFLICT 2571 2571 L -> R (IN REF. 2).
FT CONFLICT 2639 2639 Y -> S (IN REF. 3).

Query Match 75.9% Score 41; DB 1; Length 2920;
Best Local Similarity 60.0%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10
Db 1820 CVLGYGDCNC 1829

RESULT 12
EGF_PIG STANDARD; PRT; 53 AA.
AC Q00968;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Epidermal growth factor (EGF) (Fragment).
GN EGF.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Kidney;
RX MEDLINE=91197366; PubMed=2015058;
RA Pascall J.C., Jones D.S.C., Doel S.M., Clements J.M., Hunter M.,
RA Fallon T., Edwards M., Brown K.D.;
RA "Cloning and characterization of a gene encoding pig epidermal growth
RT factor.";
RL J. Mol. Endocrinol. 6:63-70(1991).
CC -!- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59516; CAA42102.1; -.
CC HSP; P01132; LEGF
CC InterPro: IPR000561; EGF-like.
CC Pfam; PF00008; EGF; 1.
CC SMART; SM00181; EGF; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 1.
CC EGF-like domain; Growth factor.
KW NON_TER 1 1
FT PEPTIDE 1 53 EPIDERMAL GROWTH FACTOR.
FT DOMAIN 2 43 EGF-LIKE.
FT DISULFID 6 20 BY SIMILARITY.
FT DISULFID 14 31 BY SIMILARITY.
```

```
FT DISULFID 33 42 BY SIMILARITY.
FT NON_TER 53 53
SQ SEQUENCE 53 AA; 6149 MW; 74F615BA4A05774D4 CRC64;

Query Match 74.1% Score 40; DB 1; Length 53;
Best Local Similarity 60.0%; Pred. No. 0.79;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10
Db 33 CVFGYVGERC 42

RESULT 13
Y402_ARCFU STANDARD; PRT; 57 AA.
ID Y402_ARCFU STANDARD; PRT; 57 AA.
AC O29845;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0402.
GN AF0402.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RP "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001076; AAB90836.1; -.
CC TYGR; AF0402; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 57 AA; 6885 MW; 4C71BDEB722568B8 CRC64;

Query Match 74.1% Score 40; DB 1; Length 57;
Best Local Similarity 50.0%; Pred. No. 0.85;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10
Db 23 CIMGYTGKRC 32

RESULT 14
NTC2_MOUSE STANDARD; PRT; 2470 AA.
ID NTC2_MOUSE STANDARD; PRT; 2470 AA.
AC O35516; Q06008;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
DE B).
GN NOTCH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RA Hamada Y., Higuchi M., Tsujimoto Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
RT single copy of mouse Notch2 gene.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RA MEDLINE=93178563; PubMed=8440332;
RX Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [3]
RN SEQUENCE OF 1765-2153 FROM N.A.
RX MEDLINE=97075110; PubMed=8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
RN [4]
RN FUNCTION.
RX MEDLINE=99396706; PubMed=10393120;
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA Tsujimoto Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT embryonic lethality.";
RL Development 126:3415-3424(1999).
RN [5]
RN DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX MEDLINE=95333893; PubMed=7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
RT mouse brain.";
RL Brain Res. Mol. Brain Res. 29:263-272(1995).
RN [6]
RN POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [7]
RN POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN [8]
RN FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (by similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.

CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC neuroepithelia, somites, optic vesicles and branchial arches, but
CC not heart.
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC the postnatal ependymal cells, and the choroid plexus throughout
CC embryonic and postnatal development.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 34, 5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D32210; BAA22094.1; -
CC EMBL; X68279; CAA48340.1; -
CC EMBL; U31881; AAC52924.1; -
CC HSSP; P16109; IFSB.
CC DR MGD; MGI:97364; Notch2.
CC DR InterPro; IPR002110; ANK.
CC DR InterPro; IPR000152; ASX_hydroxyl.
CC DR InterPro; IPR000561; EGF-like.
CC DR InterPro; IPR000742; EGF_2.
CC DR InterPro; IPR001881; EGF_CA.
CC DR InterPro; IPR001438; EGF-II.
CC DR InterPro; IPR000800; Notch.
CC DR Pfam; PF00008; EGF; 35.
CC DR Pfam; PF00023; ank; 6.
CC DR Pfam; PF00066; notch; 2.
CC DR PRINTS; PR00010; EGFBLD.
CC DR PRINTS; PR01452; NOTCH.
CC DR SMART; SM00248; ANK; 4.
CC DR SMART; SM00179; EGF_CA; 22.
CC DR SMART; SM00001; EGF_like; 12.
CC DR SMART; SM00004; NL; 3.
CC DR PROSITE; PS50088; ANK_REPEAT; 4.
CC DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC DR PROSITE; PS00010; ASX_HYDROXYL; 22.
CC DR PROSITE; PS00022; EGF_1; 33.
CC DR PROSITE; PS01186; EGF_2; 27.
CC DR PROSITE; PS01187; EGF_CA; 22.
CC DR PROSITE; PS01188; EGF_CA; 22.
CC KW Receptor; Transcription regulation; Activator; Differentiation;
CC KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
CC KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
CC KW Alternative splicing.
CC FT SIGNAL 1 25
CC FT CHAIN 26 2470
CC FT CHAIN 1666 2470
CC FT CHAIN 1697 2470
CC FT DOMAIN 26 1677
CC FT DOMAIN 1678 1698
CC FT TRANSMEM 1699 2470
CC FT DOMAIN 26 63
CC FT DOMAIN 64 102
CC FT DOMAIN 26 102

FT	DOMAIN	105	143	EGF-LIKE 3.	
FT	DOMAIN	144	180	EGF-LIKE 4.	
FT	DOMAIN	182	219	EGF-LIKE 5.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	221	256	EGF-LIKE 6.	(INCOMPLETE).
FT	DOMAIN	258	294	EGF-LIKE 7.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	296	334	EGF-LIKE 8.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	336	372	EGF-LIKE 9.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	373	411	EGF-LIKE 10.	
FT	DOMAIN	413	452	EGF-LIKE 11.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	454	490	EGF-LIKE 12.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	492	528	EGF-LIKE 13.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	530	566	EGF-LIKE 14.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	568	603	EGF-LIKE 15.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	605	641	EGF-LIKE 16.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	643	678	EGF-LIKE 17.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	680	716	EGF-LIKE 18.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	718	753	EGF-LIKE 19.	
FT	DOMAIN	755	791	EGF-LIKE 20.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	793	829	EGF-LIKE 21.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	831	869	EGF-LIKE 22.	
FT	DOMAIN	871	907	EGF-LIKE 23.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	909	945	EGF-LIKE 24.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	947	983	EGF-LIKE 25.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	985	1021	EGF-LIKE 26.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1023	1059	EGF-LIKE 27.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1061	1097	EGF-LIKE 28.	
FT	DOMAIN	1099	1145	EGF-LIKE 29.	
FT	DOMAIN	1147	1183	EGF-LIKE 30.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1185	1221	EGF-LIKE 31.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1223	1260	EGF-LIKE 32.	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1262	1300	EGF-LIKE 33.	
FT	DOMAIN	1302	1345	EGF-LIKE 34.	
FT	DOMAIN	1372	1410	EGF-LIKE 35.	
FT	REPEAT	1418	1454	LIN/NOTCH 1.	
FT	REPEAT	1501	1533	LIN/NOTCH 2.	
FT	REPEAT	1825	1869	ANK 1.	
FT	REPEAT	1874	1903	ANK 2.	
FT	REPEAT	1907	1937	ANK 3.	
FT	REPEAT	1941	1970	ANK 4.	
FT	REPEAT	1974	2003	ANK 5.	
FT	REPEAT	2007	2036	ANK 6.	
FT	DOMAIN	1645	1648	POLY-ALA.	
FT	DOMAIN	1992	1995	POLY-LEU.	
FT	DOMAIN	2183	2189	POLY-ALA.	
FT	DOMAIN	2425	2428	POLY-SER.	
FT	DOMAIN	2445	2450	POLY-GLY.	
FT	DISULFID	28	41	BY SIMILARITY.	
FT	DISULFID	35	51	BY SIMILARITY.	
FT	DISULFID	53	62	BY SIMILARITY.	
FT	DISULFID	68	79	BY SIMILARITY.	
Query Match					74.1%; Score 40; DB 1; Length 2470;
Best Local Similarity					70.0%; Pred. No. 37;
Matches					7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY	1	CVIGXSGDRC	10		
Db	324	CVWGWGSDC	333		
RESULT 15					
NTC2_HUMAN					
ID	NTC2_HUMAN	STANDARD;			
AC	Q04721; Q9H240; Q99734;				
DT	15-JUN-2002 (Rel. 41, Created)				
DT	15-JUN-2002 (Rel. 41, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Neurogenic locus notch homolog protein 2 precursor (Notch 2) (h2).				
GN	NOTCH2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				

RN
 RC
 RC
 RC
 RA
 RA
 RT
 RL
 RN
 [1]
 SEQUENCE FROM N.A.
 TISSUE=Brain;
 Blaumueller C.M., Mann R.S.;
 "Complete human notch 2 (h2) cDNA sequence.";
 Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 SEQUENCE FROM N.A.
 TISSUE=Breast tumor;
 Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.;
 "Human Notch2, a novel member of cell-fate determining NOTCH family.";
 Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE OF 967-1229 FROM N.A.
 TISSUE=Cell;
 Lemasson I., Devaux C., Mesnard J.M.;
 "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
 Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE OF 1810-2447 FROM N.A.
 TISSUE=Brain;
 MEDLINE-93265135; PubMed-1303260;
 Stifani S., Blaumueller C.M., Redhead N.J., Hill R.E.,
 Artavanis-Tsakonas S.;
 "Human homologs of a Drosophila enhancer of split gene product define a novel family of nuclear proteins.";
 Nat. Genet. 2:119-127(1992).
 [5]
 POST-TRANSLATIONAL PROCESSING.
 MEDLINE-97386453; PubMed-9244302;
 Blaumueller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
 "Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane.";
 Cell 90:281-291(1997).
 [6]
 IDENTIFICATION OF LIGANDS.
 MEDLINE-99180765; PubMed-10079256;
 Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L., Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 "Human ligands of the Notch receptor.";
 Am. J. Pathol. 154:785-794(1999).
 -!- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs (By similarity).
 -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).
 -!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.
 -!- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung, skeletal muscle and liver.
 -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active ligand-accessible form. Cleavage results in a C-terminal fragment N(TM) and a N-terminal fragment N(EC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).
 -!- PTM: Phosphorylated (By similarity).
 -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 -!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
 -!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 -!- SIMILARITY: CONTAINS 6 ANK REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; AF308601; AAA36377.2; --
 CC EMBL; AF315356; AAG37073.1; --
 CC EMBL; U77493; AAB19224.1; --
 CC HSP; P00740; IEDM.
 CC Genew; HGNC:7882; NOTCH2.
 CC MIM; 600275; --
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001891; EGF_Ca.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR000800; Notch.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00066; notch; 2.
 CC PRINTS; PRO0010; EGFBLD.
 CC PRINTS; PRO0011; EGFLAMIN.
 CC SMART; SM00248; ANK; 4.
 CC SMART; SM00179; EGF_CA; 22.
 CC SMART; SM00001; EGF_like; 12.
 CC SMART; SM00004; NL; 2.
 CC PROSITE; PS50088; ANK_REPEAT; 4.
 CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE; PS00010; ASX_HYDROXYL; 22.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 29.
 CC PROSITE; PS01187; EGF_CA; 22.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 2471
 FT CHAIN 1686 2471
 FT CHAIN 1697 2471
 FT DOMAIN 26 1677
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2471
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 180
 FT DOMAIN 182 219
 FT DOMAIN 221 258
 FT DOMAIN 260 296
 FT DOMAIN 298 336
 FT DOMAIN 338 374
 FT DOMAIN 375 413
 FT DOMAIN 415 454
 FT DOMAIN 456 492
 FT DOMAIN 494 530
 FT DOMAIN 532 568
 FT DOMAIN 570 605
 FT DOMAIN 607 643
 FT DOMAIN 645 680
 FT DOMAIN 682 718
 FT DOMAIN 720 755
 FT DOMAIN 757 793
 FT DOMAIN 795 831
 FT DOMAIN 833 871
 FT DOMAIN 873 909
 FT DOMAIN 911 947
 FT DOMAIN 949 985
 FT DOMAIN 987 1023
 FT DOMAIN 1025 1061
 FT DOMAIN 1063 1099
 FT DOMAIN 1101 1147
 FT DOMAIN 1149 1185
 FT DOMAIN 1187 1223
 FT DOMAIN 1225 1262
 FT DOMAIN 1264 1302
 FT DOMAIN 1304 1343
 FT DOMAIN 1374 1412
 FT REPEAT 1420 1456
 FT REPEAT 1456 1503
 FT REPEAT 1503 1535
 FT REPEAT 1535 1571
 FT REPEAT 1571 1607
 FT REPEAT 1607 1643
 FT REPEAT 1643 1679
 FT REPEAT 1679 1715
 FT REPEAT 1715 1751
 FT REPEAT 1751 1787
 FT REPEAT 1787 1823
 FT REPEAT 1823 1859
 FT REPEAT 1859 1895
 FT REPEAT 1895 1931
 FT REPEAT 1931 1967
 FT REPEAT 1967 2003
 FT REPEAT 2003 2039
 FT REPEAT 2039 2075
 FT REPEAT 2075 2111
 FT REPEAT 2111 2147
 FT REPEAT 2147 2183
 FT REPEAT 2183 2219
 FT REPEAT 2219 2255
 FT REPEAT 2255 2291
 FT REPEAT 2291 2327
 FT REPEAT 2327 2363
 FT REPEAT 2363 2399
 FT REPEAT 2399 2435
 FT REPEAT 2435 2471
 FT REPEAT 2471 2507
 FT REPEAT 2507 2543
 FT REPEAT 2543 2579
 FT REPEAT 2579 2615
 FT REPEAT 2615 2651
 FT REPEAT 2651 2687
 FT REPEAT 2687 2723
 FT REPEAT 2723 2759
 FT REPEAT 2759 2795
 FT REPEAT 2795 2831
 FT REPEAT 2831 2867
 FT REPEAT 2867 2903
 FT REPEAT 2903 2939
 FT REPEAT 2939 2975
 FT REPEAT 2975 3011
 FT REPEAT 3011 3047
 FT REPEAT 3047 3083
 FT REPEAT 3083 3119
 FT REPEAT 3119 3155
 FT REPEAT 3155 3191
 FT REPEAT 3191 3227
 FT REPEAT 3227 3263
 FT REPEAT 3263 3299
 FT REPEAT 3299 3335
 FT REPEAT 3335 3371
 FT REPEAT 3371 3407
 FT REPEAT 3407 3443
 FT REPEAT 3443 3479
 FT REPEAT 3479 3515
 FT REPEAT 3515 3551
 FT REPEAT 3551 3587
 FT REPEAT 3587 3623
 FT REPEAT 3623 3659
 FT REPEAT 3659 3695
 FT REPEAT 3695 3731
 FT REPEAT 3731 3767
 FT REPEAT 3767 3803
 FT REPEAT 3803 3839
 FT REPEAT 3839 3875
 FT REPEAT 3875 3911
 FT REPEAT 3911 3947
 FT REPEAT 3947 3983
 FT REPEAT 3983 4019
 FT REPEAT 4019 4055
 FT REPEAT 4055 4091
 FT REPEAT 4091 4127
 FT REPEAT 4127 4163
 FT REPEAT 4163 4199
 FT REPEAT 4199 4235
 FT REPEAT 4235 4271
 FT REPEAT 4271 4307
 FT REPEAT 4307 4343
 FT REPEAT 4343 4379
 FT REPEAT 4379 4415
 FT REPEAT 4415 4451
 FT REPEAT 4451 4487
 FT REPEAT 4487 4523
 FT REPEAT 4523 4559
 FT REPEAT 4559 4595
 FT REPEAT 4595 4631
 FT REPEAT 4631 4667
 FT REPEAT 4667 4703
 FT REPEAT 4703 4739
 FT REPEAT 4739 4775
 FT REPEAT 4775 4811
 FT REPEAT 4811 4847
 FT REPEAT 4847 4883
 FT REPEAT 4883 4919
 FT REPEAT 4919 4955
 FT REPEAT 4955 4991
 FT REPEAT 4991 5027
 FT REPEAT 5027 5063
 FT REPEAT 5063 5099
 FT REPEAT 5099 5135
 FT REPEAT 5135 5171
 FT REPEAT 5171 5207
 FT REPEAT 5207 5243
 FT REPEAT 5243 5279
 FT REPEAT 5279 5315
 FT REPEAT 5315 5351
 FT REPEAT 5351 5387
 FT REPEAT 5387 5423
 FT REPEAT 5423 5459
 FT REPEAT 5459 5495
 FT REPEAT 5495 5531
 FT REPEAT 5531 5567
 FT REPEAT 5567 5603
 FT REPEAT 5603 5639
 FT REPEAT 5639 5675
 FT REPEAT 5675 5711
 FT REPEAT 5711 5747
 FT REPEAT 5747 5783
 FT REPEAT 5783 5819
 FT REPEAT 5819 5855
 FT REPEAT 5855 5891
 FT REPEAT 5891 5927
 FT REPEAT 5927 5963
 FT REPEAT 5963 6000
 FT REPEAT 6000 6036
 FT REPEAT 6036 6072
 FT REPEAT 6072 6108
 FT REPEAT 6108 6144
 FT REPEAT 6144 6180
 FT REPEAT 6180 6216
 FT REPEAT 6216 6252
 FT REPEAT 6252 6288
 FT REPEAT 6288 6324
 FT REPEAT 6324 6360
 FT REPEAT 6360 6396
 FT REPEAT 6396 6432
 FT REPEAT 6432 6468
 FT REPEAT 6468 6504
 FT REPEAT 6504 6540
 FT REPEAT 6540 6576
 FT REPEAT 6576 6612
 FT REPEAT 6612 6648
 FT REPEAT 6648 6684
 FT REPEAT 6684 6720
 FT REPEAT 6720 6756
 FT REPEAT 6756 6792
 FT REPEAT 6792 6828
 FT REPEAT 6828 6864
 FT REPEAT 6864 6900
 FT REPEAT 6900 6936
 FT REPEAT 6936 6972
 FT REPEAT 6972 7008
 FT REPEAT 7008 7044
 FT REPEAT 7044 7080
 FT REPEAT 7080 7116
 FT REPEAT 7116 7152
 FT REPEAT 7152 7188
 FT REPEAT 7188 7224
 FT REPEAT 7224 7260
 FT REPEAT 7260 7296
 FT REPEAT 7296 7332
 FT REPEAT 7332 7368
 FT REPEAT 7368 7404
 FT REPEAT 7404 7440
 FT REPEAT 7440 7476
 FT REPEAT 7476 7512
 FT REPEAT 7512 7548
 FT REPEAT 7548 7584
 FT REPEAT 7584 7620
 FT REPEAT 7620 7656
 FT REPEAT 7656 7692
 FT REPEAT 7692 7728
 FT REPEAT 7728 7764
 FT REPEAT 7764 7800
 FT REPEAT 7800 7836
 FT REPEAT 7836 7872
 FT REPEAT 7872 7908
 FT REPEAT 7908 7944
 FT REPEAT 7944 7980
 FT REPEAT 7980 8016
 FT REPEAT 8016 8052
 FT REPEAT 8052 8088
 FT REPEAT 8088 8124
 FT REPEAT 8124 8160
 FT REPEAT 8160 8196
 FT REPEAT 8196 8232
 FT REPEAT 8232 8268
 FT REPEAT 8268 8304
 FT REPEAT 8304 8340
 FT REPEAT 8340 8376
 FT REPEAT 8376 8412
 FT REPEAT 8412 8448
 FT REPEAT 8448 8484
 FT REPEAT 8484 8520
 FT REPEAT 8520 8556
 FT REPEAT 8556 8592
 FT REPEAT 8592 8628
 FT REPEAT 8628 8664
 FT REPEAT 8664 8700
 FT REPEAT 8700 8736
 FT REPEAT 8736 8772
 FT REPEAT 8772 8808
 FT REPEAT 8808 8844
 FT REPEAT 8844 8880
 FT REPEAT 8880 8916
 FT REPEAT 8916 8952
 FT REPEAT 8952 8988
 FT REPEAT 8988 9024
 FT REPEAT 9024 9060
 FT REPEAT 9060 9096
 FT REPEAT 9096 9132
 FT REPEAT 9132 9168
 FT REPEAT 9168 9204
 FT REPEAT 9204 9240
 FT REPEAT 9240 9276
 FT REPEAT 9276 9312
 FT REPEAT 9312 9348
 FT REPEAT 9348 9384
 FT REPEAT 9384 9420
 FT REPEAT 9420 9456
 FT REPEAT 9456 9492
 FT REPEAT 9492 9528
 FT REPEAT 9528 9564
 FT REPEAT 9564 9600
 FT REPEAT 9600 9636
 FT REPEAT 9636 9672
 FT REPEAT 9672 9708
 FT REPEAT 9708 9744
 FT REPEAT 9744 9780
 FT REPEAT 9780 9816
 FT REPEAT 9816 9852
 FT REPEAT 9852 9888
 FT REPEAT 9888 9924
 FT REPEAT 9924 9960
 FT REPEAT 9960 10000
 FT REPEAT 10000 10040
 FT REPEAT 10040 10080
 FT REPEAT 10080 10120
 FT REPEAT 10120 10160
 FT REPEAT 10160 10200
 FT REPEAT 10200 10240
 FT REPEAT 10240 10280
 FT REPEAT 10280 10320
 FT REPEAT 10320 10360
 FT REPEAT 10360 10400
 FT REPEAT 10400 10440
 FT REPEAT 10440 10480
 FT REPEAT 10480 10520
 FT REPEAT 10520 10560
 FT REPEAT 10560 10600
 FT REPEAT 10600 10640
 FT REPEAT 10640 10680
 FT REPEAT 10680 10720
 FT REPEAT 10720 10760
 FT REPEAT 10760 10800
 FT REPEAT 10800 10840
 FT REPEAT 10840 10880
 FT REPEAT 10880 10920
 FT REPEAT 10920 10960
 FT REPEAT 10960 11000
 FT REPEAT 11000 11040
 FT REPEAT 11040 11080
 FT REPEAT 11080 11120
 FT REPEAT 11120 11160
 FT REPEAT 11160 11200
 FT REPEAT 11200 11240
 FT REPEAT 11240 11280
 FT REPEAT 11280 11320
 FT REPEAT 11320 11360
 FT REPEAT 11360 11400
 FT REPEAT 11400 11440
 FT REPEAT 11440 11480
 FT REPEAT 11480 11520
 FT REPEAT 11520 11560
 FT REPEAT 11560 11600
 FT REPEAT 11600 11640
 FT REPEAT 11640 11680
 FT REPEAT 11680 11720
 FT REPEAT 11720 11760
 FT REPEAT 11760 11800
 FT REPEAT 11800 11840
 FT REPEAT 11840 11880
 FT REPEAT 11880 11920
 FT REPEAT 11920 11960
 FT REPEAT 11960 12000
 FT REPEAT 12000 12040
 FT REPEAT 12040 12080
 FT REPEAT 12080 12120
 FT REPEAT 12120 12160
 FT REPEAT 12160 12200
 FT REPEAT 12200 12240
 FT REPEAT 12240 12280
 FT REPEAT 12280 12320
 FT REPEAT 12320 12360
 FT REPEAT 12360 12400
 FT REPEAT 12400 12440
 FT REPEAT 12440 12480
 FT REPEAT 12480 12520
 FT REPEAT 12520 12560
 FT REPEAT 12560 12600
 FT REPEAT 12600 12640
 FT REPEAT 12640 12680
 FT REPEAT 12680 12720
 FT REPEAT 12720 12760
 FT REPEAT 12760 12800
 FT REPEAT 12800 12840
 FT REPEAT 12840 12880
 FT REPEAT 12880 12920
 FT REPEAT 12920 12960
 FT REPEAT 12960 13000
 FT REPEAT 13000 13040
 FT REPEAT 13040 13080
 FT REPEAT 13080 13120
 FT REPEAT 13120 13160
 FT REPEAT 13160 13200
 FT REPEAT 13200 13240
 FT REPEAT 13240 13280
 FT REPEAT 13280 13320
 FT REPEAT 13320 13360
 FT REPEAT 13360 13400
 FT REPEAT 13400 13440
 FT REPEAT 13440 13480
 FT REPEAT 13480 13520
 FT REPEAT 13520 13560
 FT REPEAT 13560 13600
 FT REPEAT 13600 13640
 FT REPEAT 13640 13680
 FT REPEAT 13680 13720
 FT REPEAT 13720 13760
 FT REPEAT 13760 13800
 FT REPEAT 13800 13840
 FT REPEAT 13840 13880
 FT REPEAT 13880 13920
 FT REPEAT 13920 13960
 FT REPEAT 13960 14000
 FT REPEAT 14000 14040
 FT REPEAT 14040 14080
 FT REPEAT 14080 14120
 FT REPEAT 14120 14160
 FT REPEAT 14160 14200
 FT REPEAT 14200 14240
 FT REPEAT 14240 14280
 FT REPEAT 14280 14320
 FT REPEAT 14320 14360
 FT REPEAT 14360 14400
 FT REPEAT 14400 14440
 FT REPEAT 14440 14480
 FT REPEAT 14480 14520
 FT REPEAT 14520 14560
 FT REPEAT 14560 14600
 FT REPEAT 14600 14640
 FT REPEAT 14640 14680
 FT REPEAT 14680 14720
 FT REPEAT 14720 14760
 FT REPEAT 14760 14800
 FT REPEAT 14800 14840
 FT REPEAT 14840 14880
 FT REPEAT 14880 14920
 FT REPEAT 14920 14960
 FT REPEAT 14960 15000
 FT REPEAT 15000 15040
 FT REPEAT 15040 15080
 FT REPEAT 15080 15120
 FT REPEAT 15120 15160
 FT REPEAT 15160 15200
 FT REPEAT 15200 15240
 FT REPEAT 15240 15280
 FT REPEAT 15280 15320
 FT REPEAT 15320 15360
 FT REPEAT 15360 15400
 FT REPEAT 15400 15440
 FT REPEAT 15440 15480
 FT REPEAT 15480 15520
 FT REPEAT 15520 15560
 FT REPEAT 15560 15600
 FT REPEAT 15600 15640
 FT REPEAT 15640 15680
 FT REPEAT 15680 15720
 FT REPEAT 15720 15760
 FT REPEAT 15760 15800
 FT REPEAT 15800 15840
 FT REPEAT 15840 15880
 FT REPEAT 15880 15920
 FT REPEAT 15920 15960
 FT REPEAT 15960 16000
 FT REPEAT 16000 16040
 FT REPEAT 16040 16080
 FT REPEAT 16080 16120
 FT REPEAT 16120 16160
 FT REPEAT 16160 16200
 FT REPEAT 16200 16240
 FT REPEAT 16240 16280
 FT REPEAT 16280 16320
 FT REPEAT 16320 16360
 FT REPEAT 16360 16400
 FT REPEAT 16400 16440
 FT REPEAT 16440 16480
 FT REPEAT 16480 16520
 FT REPEAT 16520 16560
 FT REPEAT 16560 16600
 FT REPEAT 16600 16640
 FT REPEAT 16640 16680
 FT REPEAT 16680 16720
 FT REPEAT 16720 16760
 FT REPEAT 16760 16800
 FT REPEAT 16800 16840
 FT REPEAT 16840 16880
 FT REPEAT 16880 16920
 FT REPEAT 16920 16960
 FT REPEAT 16960 17000
 FT REPEAT 17000 17040
 FT REPEAT 17040 17080
 FT REPEAT 17080 17120
 FT REPEAT 17120 17160
 FT REPEAT 17160 17200
 FT REPEAT 17200 17240
 FT REPEAT 17240 17280
 FT REPEAT 17280 17320
 FT REPEAT 17320 17360
 FT REPEAT 17360 17400
 FT REPEAT 17400 17440
 FT REPEAT 17440 17480
 FT REPEAT 17480 17520
 FT REPEAT 17520 17560
 FT REPEAT 17560 17600
 FT REPEAT 17600 17640
 FT REPEAT 17640 17680
 FT REPEAT 17680 17720
 FT REPEAT 17720 17760
 FT REPEAT 17760 17800
 FT REPEAT 17800 17840
 FT REPEAT 17840 17880
 FT REPEAT 17880 17920
 FT REPEAT 17920 17960
 FT REPEAT 17960 18000
 FT REPEAT 18000 18040
 FT REPEAT 18040 18080
 FT REPEAT 18080 18120
 FT REPEAT 18120 18160
 FT REPEAT 18160 18200
 FT REPEAT 18200 18240
 FT REPEAT 18240 18280
 FT REPEAT 18280 18320
 FT REPEAT 18320 18360
 FT REPEAT 18360 18400
 FT REPEAT 18400 18440
 FT REPEAT 18440 18480
 FT REPEAT 18480 18520
 FT REPEAT 18520 18560
 FT REPEAT 18560 18600
 FT REPEAT 18600 18640
 FT REPEAT 18640 18680
 FT REPEAT 18680 18720
 FT REPEAT 18720 18760
 FT REPEAT 18760 18800
 FT REPEAT 18800 18840
 FT REPEAT 18840 18880
 FT REPEAT 18880 18920
 FT REPEAT 18920 18960
 FT REPEAT 18960 19000
 FT REPEAT 19000 19040
 FT REPEAT 19040 19080
 FT REPEAT 19080 19120
 FT REPEAT 19120 19160
 FT REPEAT 19160 19200
 FT REPEAT 19200 19240
 FT REPEAT 19240 19280
 FT REPEAT 19280 19320
 FT REPEAT 19320 19360
 FT REPEAT 19360 19400
 FT REPEAT 19400 19440
 FT REPEAT 19440 19480
 FT REPEAT 19480 19520
 FT REPEAT 19520 19560
 FT REPEAT 19560 19600
 FT REPEAT 19600 19640
 FT REPEAT 19640 19680
 FT REPEAT 19680 19720
 FT REPEAT 19720 19760
 FT REPEAT 19760 19800
 FT REPEAT 19800 19840
 FT REPEAT 19840 19880
 FT REPEAT 19880 19920
 FT REPEAT 19920 19960
 FT REPEAT 19960 20000
 FT REPEAT 20000 20040
 FT REPEAT 20040 20080
 FT REPEAT 20080 20120
 FT REPEAT 20120 20160
 FT REPEAT 20160 20200
 FT REPEAT 20200 20240
 FT REPEAT 20240 20280
 FT REPEAT 20280 20320
 FT REPEAT 20320 20360
 FT REPEAT 20360 20400
 FT REPEAT 20400 20440
 FT REPEAT 20440 20480
 FT REPEAT 20480 20520
 FT REPEAT 20520 20560
 FT REPEAT 20560 20600
 FT REPEAT 20600 20640
 FT REPEAT 20640 20680
 FT REPEAT 20680 20720
 FT REPEAT 20720 20760
 FT REPEAT 20760 20800
 FT REPEAT 20800 20840
 FT REPEAT 20840 20880
 FT REPEAT 20880 20920
 FT REPEAT 20920 20960
 FT REPEAT 20960 21000
 FT REPEAT 21000 21040
 FT REPEAT 21040 21080
 FT REPEAT 21080 21120
 FT REPEAT 21120 21160
 FT REPEAT 21160 21200
 FT REPEAT 21200 21240
 FT REPEAT 21240 21280
 FT REPEAT 21280 21320
 FT REPEAT 21320 21360
 FT REPEAT 21360 21400
 FT REPEAT 21400 21440
 FT REPEAT 21440 21480
 FT REPEAT 21480 21520
 FT REPEAT 21520 21560
 FT REPEAT 21560 21600
 FT REPEAT 21600 21640
 FT REPEAT 21640 21680
 FT REPEAT 21680 21720
 FT REPEAT 21720 21760
 FT REPEAT 21760 21800
 FT REPEAT 21800 21840
 FT REPEAT 21840 21880
 FT REPEAT 21880 21920
 FT REPEAT 21920 21960
 FT REPEAT 21960 22000
 FT REPEAT 22000 22040
 FT REPEAT 22040 22080
 FT REPEAT 22080 22120
 FT REPEAT 22120 22160
 FT REPEAT 22160 22200
 FT REPEAT 22200 22240
 FT REPEAT 22240 22280
 FT REPEAT 22280 22320
 FT REPEAT 22320 22360
 FT REPEAT 22360 22400
 FT REPEAT 22400 22440
 FT REPEAT 22440 22480
 FT REPEAT 22480 22520
 FT REPEAT 22520 22560
 FT REPEAT 22560 22600
 FT REPEAT 22600 22640
 FT REPEAT 22640 22680
 FT REPEAT 22680 22720
 FT REPEAT 22720 22760
 FT REPEAT 22760 22800
 FT REPEAT 22800 22840
 FT REPEAT 22840 22880
 FT REPEAT 22880 22920
 FT REPEAT 22920 22960
 FT REPEAT 22960 23000
 FT REPEAT 23000 23040
 FT REPEAT 23040 23080
 FT REPEAT 23080 23120
 FT REPEAT 23120 23160
 FT REPEAT 23160 23200
 FT REPEAT 23200 23240
 FT REPEAT 23240 23280
 FT REPEAT 23280 23320
 FT REPEAT 23320 23360
 FT REPEAT 23360 23400
 FT REPEAT 23400 23440
 FT REPEAT 23440 23480
 FT REPEAT 23480 23520
 FT REPEAT 23520 23560
 FT REPEAT 23560 23600
 FT REPEAT 23600 23640
 FT REPEAT 23640 23680
 FT REPEAT 23680 23720
 FT REPEAT 23720 23760
 FT REPEAT 23760 23800
 FT REPEAT 23800 23840
 FT REPEAT 23840 23880
 FT REPEAT 23880 23920
 FT REPEAT 23920 23960
 FT REPEAT 23960 24000
 FT REPEAT 24000 24040
 FT REPEAT 24040 24080
 FT

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 41.333 Seconds
(without alignments)
49.850 Million cell updates/sec

Title: US-09-673-785D-8
Perfect score: 54
Sequence: 1 CVIGXSGDRC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rhodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	1200	11 Q8VD07	Q8vd07 mus musculus
2	49	90.7	2146	5 Q9VC97	Q9vc97 drosophila
3	44	81.5	106	4 Q9UN93	Q9un93 homo sapien
4	44	81.5	308	4 Q9UN94	Q9un94 homo sapien
5	44	81.5	366	4 Q9UN95	Q9un95 homo sapien
6	44	81.5	372	5 Q21756	Q21756 caenorhabdi
7	44	81.5	877	4 Q9H3Q6	Q9h3q6 homo sapien
8	44	81.5	878	4 Q9G2Z2	Q9gz22 homo sapien
9	44	81.5	878	4 Q9H3Q7	Q9h3q7 homo sapien
10	44	81.5	901	4 Q9H195	Q9h195 homo sapien
11	44	81.5	957	4 Q14651	Q14651 homo sapien
12	44	81.5	1217	4 Q9UKW9	Q9ukw9 homo sapien
13	43	79.6	89	6 Q28867	Q28867 equus cabal
14	43	79.6	264	5 Q20043	Q20043 caenorhabdi
15	42	77.8	165	13 Q90XX0	Q90xx0 orectolobus
16	42	77.8	1162	5 Q8WTP0	Q8wtp0 halocynthia

17	42	77.8	2352	5	O61240	O61240 halocynthia
18	42	77.8	2531	5	O16004	O16004 lytechinus
19	42	77.8	4006	11	O35452	O35452 mus musculus
20	42	77.8	4114	11	O54796	O54796 mus musculus
21	41	75.9	379	11	O35883	O35883 rattus norv
22	41	75.9	447	11	O63348	O63348 rattus norv
23	41	75.9	601	10	O9LYF1	O9lyf1 arabidopsis
24	41	75.9	1531	11	O9WVB5	O9wvb5 mus musculus
25	41	75.9	2920	11	O9ROM0	O9rom0 mus musculus
26	41	75.9	2972	5	P90891	P90891 caenorhabdi
27	40	74.1	601	5	Q22913	Q22913 caenorhabdi
28	40	74.1	708	5	Q9NFS9	Q9nfs9 drosophila
29	40	74.1	969	4	Q96KG6	Q96kg6 homo sapien
30	40	74.1	1035	5	Q9NEG1	Q9neg1 drosophila
31	40	74.1	1210	6	Q95ND4	Q95nd4 felis silve
32	40	74.1	1214	6	Q9BDQ0	Q9bdq0 sus scrofa
33	40	74.1	1216	6	Q9BEA0	Q9bea0 canis famil
34	40	74.1	2524	5	Q9GPA5	Q9gpa5 branchiosto
35	40	74.1	4072	5	Q9WAY4	Q9way4 drosophila
36	39	72.2	191	5	Q8T5Z1	Q8t5z1 caenorhabdi
37	39	72.2	191	5	Q8T5Z0	Q8t5z0 caenorhabdi
38	39	72.2	193	5	Q8T5Z2	Q8t5z2 caenorhabdi
39	39	72.2	193	5	Q8STG0	Q8stg0 caenorhabdi
40	39	72.2	194	5	Q8T5Z3	Q8t5z3 caenorhabdi
41	39	72.2	473	4	Q9BWZ9	Q9bwz9 homo sapien
42	39	72.2	473	4	Q12891	Q12891 homo sapien
43	39	72.2	615	13	O57409	O57409 brachydanio
44	39	72.2	963	5	Q9GPM9	Q9gpm9 caenorhabdi
45	39	72.2	1045	5	Q8T3A6	Q8t3a6 caenorhabdi

ALIGNMENTS

RESULT 1

Q8VD07 ID Q8VD07 PRELIMINARY; PRT; 1200 AA.
AC Q8VD07;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to epidermal growth factor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017681; AAHL7681.1; -;
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001336; EGF_1.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 8.
DR Pfam; PF00058; ldl_recept_b; 7.
DR PRINTS; PRO0009; EGFTGF.
DR SMART; SM00181; EGF; 9.
DR SMART; SM00179; EGF_CA; 8.
DR SMART; SM00135; LV; 9.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_6.
DR PROSITE; PS01187; EGF_CA; UNKNOWN_3.
SQ SEQUENCE 1200 AA; 131317 MW; B6A44F7294746476 CRC64;

Query Match 100.0%; Score 54; DB 11; Length 1200;

Best Local Similarity 90.0%; Pred. No. 0.13; Indels 0; Gaps 0;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CVIGXSGDRC 10
DB 992 CVIGXSGDRC 1001

RESULT 2
ID Q9VC97 PRELIMINARY; PRT; 2146 AA.
AC Q9VC97;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE CRB protein.
DE CRB OR CG6383.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley B.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AF003747; AAF56276.1; -.
DR HSSP; P00740; 1EDM.
DR FlyBase; FBgn0000368; crb.
DR InterPro; IPR000152; Asx-hydroxyl.
DR InterPro; IPR0000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF000008; EGF; 26.
DR Pfam; PF00054; laminin_G; 3.

```

```

DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFELAMININ.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00001; EGF_like; 16.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 16.
DR PROSITE; PS00022; EGF_1; 25.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 13.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2146 AA; 233570 MW; 8E2B9E32B761115 CRC64;

Query Match 90.7%; Score 49; DB 5; Length 2146;
Best Local Similarity 70.0%; Pred. NO. 2.2;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10
   I : : : : :
Db 569 CAVIGSGDRC 578

RESULT 3
Q9UN93 PRELIMINARY; PRT; 106 AA.
ID Q9UN93;
AC Q9UN93;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Mucin 3 (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99333563; PubMed=10405327;
RA Williams S.J., Munster D.J., Quin R.J., Gotley D.C., McGuckin M.A.;
RT "The MUC3 gene encodes a transmembrane mucin and is alternatively
   spliced."
RL Biochem. Biophys. Res. Commun. 261:83-89(1999).
DR EMBL; AF143373; AAD45884.1; -.
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 106 AA; 11340 MW; 5A2E3C24C905E182 CRC64;

Query Match 81.5%; Score 44; DB 4; Length 106;
Best Local Similarity 70.0%; Pred. NO. 0.88;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDRC 10
   I : : : : :
Db 14 CLPFGSGDRC 23

RESULT 4
Q9UN94 PRELIMINARY; PRT; 308 AA.
ID Q9UN94;
AC Q9UN94;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Mucin 3 (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN  SEQUENCE FROM N.A.
RP  TISSUE=COLON MUCOSA;
RX  MEDLINE=99335363; PubMed=10405327;
RA  Williams S.J., Munster D.J., Quin R.J., Gotley D.C., McGuckin M.A.;
RT  "The MUC3 gene encodes a transmembrane mucin and is alternatively
RL  spliced.";
RL  Biochem. Biophys. Res. Commun. 261:83-89(1999).
DR  EMBL; AF143372; AAD45883.1; -.
DR  InterPro; IPR000561; EGF-like.
DR  InterPro; IPR000082; SEA_domain.
DR  Pfam; PF01390; SEA; 1.
DR  SMART; SM00181; EGF; 1.
DR  SMART; SM00200; SEA; 1.
DR  PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR  PROSITE; PS01186; EGF_2; 1.
DR  PROSITE; PS00024; SEA; 1.
KW  EGF-like domain; Glycoprotein.
FT  NON_TER 1
SQ  SEQUENCE 308 AA; 34665 MW; A5782A702D4EAF86 CRC64;

Query Match      81.5%; Score 44; DB 4; Length 308;
Best Local Similarity 70.0%; Pred. No. 2.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY  1 CVIGXSGDRC 10
Db  14 CLPFGSGDRC 23

RESULT 5
Q9UN95  PRELIMINARY; PRT; 366 AA.
AC  Q9UN95;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAY-2000 (TrEMBLrel. 20, Last annotation update)
DE  Mucin 3 (Fragment).
GN  MUC3.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  TISSUE=COLON MUCOSA;
RA  Williams S.J., Munster D.J., Quin R.J., Gotley D.C., McGuckin M.A.;
RT  "The MUC3 gene encodes a transmembrane mucin and is alternatively
RL  spliced.";
RL  Biochem. Biophys. Res. Commun. 261:83-89(1999).
DR  EMBL; AF143371; AAD45882.1; -.
DR  InterPro; IPR000561; EGF-like.
DR  InterPro; IPR000082; SEA_domain.
DR  Pfam; PF01390; SEA; 1.
DR  SMART; SM00181; EGF; 1.
DR  SMART; SM00200; SEA; 1.
DR  PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR  PROSITE; PS01186; EGF_2; 1.
DR  PROSITE; PS00024; SEA; 1.
KW  EGF-like domain; Glycoprotein.
FT  NON_TER 1
SQ  SEQUENCE 366 AA; 40953 MW; 88F3F0E3F439A3C2 CRC64;

Query Match      81.5%; Score 44; DB 4; Length 366;
Best Local Similarity 70.0%; Pred. No. 3.2;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY  1 CVIGXSGDRC 10
Db  14 CLPFGSGDRC 23

RESULT 6
Q21756  PRELIMINARY; PRT; 372 AA.
AC  Q21756;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  Hypothetical 39.1 kDa protein.
GN  R05G6.9.
OS  Caenorhabditis elegans.
OC  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC  Rhabditidae; Peloderinae; Caenorhabditis.
OX  NCBI_TaxID=6239;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN-BRISTOL N2;
RX  MEDLINE=99069613; PubMed=9851916;
RA  None;
RT  "Genome sequence of the nematode C. elegans: a platform for
RL  investigating biology. The C. elegans Sequencing Consortium.";
RL  Science 282:2012-2018(1998).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  STRAIN-BRISTOL N2;
RA  Murray J., Le T.T.;
RT  "The sequence of C. elegans cosmid R05G6.";
RL  Submitted (MAY-1996) to the EMBL/GenBank/DBDJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RX  STRAIN-BRISTOL N2;
RA  Waterston R.;
RT  "Direct Submission.";
RL  EMBL; U58746; AAB00626.1; -.
DR  HSSP; P01132; 1A3P.
DR  InterPro; IPR000561; EGF-like.
DR  Pfam; PF00008; EGF; 5.
DR  SMART; SM00181; EGF; 7.
DR  PROSITE; PS00022; EGF_1; UNKNOWN_7.
DR  PROSITE; PS01186; EGF_2; 6.
KW  EGF-like domain; Glycoprotein; Hypothetical protein.
SQ  SEQUENCE 372 AA; 39085 MW; DB36AB251EEB6884 CRC64;

Query Match      81.5%; Score 44; DB 5; Length 372;
Best Local Similarity 60.0%; Pred. No. 3.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY  1 CVIGXSGDRC 10
Db  233 CYLGSGDKC 242

RESULT 7
Q9H3Q6  PRELIMINARY; PRT; 877 AA.
AC  Q9H3Q6;
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Intestinal mucin (Fragment).
GN  MUC3A.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  TISSUE=SMALL INTESTINE;
RX  MEDLINE=21183349; PubMed=11289722;
RA  Kyo K., Muto T., Nagawa H., Lathrop GM., Nakamura Y.;
RT  "Associations of distinct variants of the intestinal mucin gene MUC3A
RL  with ulcerative colitis and Crohn's disease.";
RL  J. Hum. Genet. 46:5-20(2001).
```

```
DR EMBL; AB038784; BAB12118.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 877 AA; 93720 MW; 09EE5B86BC838DFA CRC64;

Query Match      81.5%; Score 44; DB 4; Length 877;
Best Local Similarity 70.0%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 525 CLPGFSGDRC 534
I: |:|||||

RESULT 8
ID Q9GZ22 PRELIMINARY; PRT; 878 AA.
AC Q9GZ22;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Intestinal mucin (Fragment).
GN MUC3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kyo K., Parkes M., Ishigami H., Muto T., Nagawa H., Nakamura Y.;
RT "Human intestinal mucin gene 'MUC3' consists of two genes, MUC3A and
RT MUC3B: Associations of distinct variants of MUC3B with ulcerative
RT colitis and Crohn's disease.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038783; BAB12117.1; -.
DR EMBL; AB038781; BAB12115.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 878 AA; 93685 MW; 1414E7E00AF5B4C CRC64;

Query Match      81.5%; Score 44; DB 4; Length 878;
Best Local Similarity 70.0%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 526 CLPGFSGDRC 535
I: |:|||||

RESULT 9
ID Q9H3Q7 PRELIMINARY; PRT; 878 AA.
AC Q9H3Q7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Intestinal mucin (Fragment).
GN MUC3A.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kyo K., Parkes M., Ishigami H., Muto T., Nagawa H., Nakamura Y.;
RT "Human intestinal mucin gene 'MUC3' consists of two genes, MUC3A and
RT MUC3B: Associations of distinct variants of MUC3B with ulcerative
RT colitis and Crohn's disease.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038783; BAB12117.1; -.
DR EMBL; AB038781; BAB12115.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 878 AA; 93685 MW; 1414E7E00AF5B4C CRC64;

Query Match      81.5%; Score 44; DB 4; Length 878;
Best Local Similarity 70.0%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 526 CLPGFSGDRC 535
I: |:|||||

RESULT 9
ID Q9H3Q7 PRELIMINARY; PRT; 878 AA.
AC Q9H3Q7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Intestinal mucin (Fragment).
GN MUC3A.
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21183349; PubMed=11289722;
RA Kyo K., Muto T., Nagawa H., Lathrop GM., Nakamura Y.;
RT "Associations of distinct variants of the intestinal mucin gene MUC3A
RT with ulcerative colitis and Crohn's disease.";
RL J. Hum. Genet. 46:5-20(2001).
DR EMBL; AB038782; BAB12116.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 878 AA; 93824 MW; B36E9B97F1EC5795 CRC64;

Query Match      81.5%; Score 44; DB 4; Length 878;
Best Local Similarity 70.0%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 526 CLPGFSGDRC 535
I: |:|||||

RESULT 10
ID Q9H195 PRELIMINARY; PRT; 901 AA.
AC Q9H195;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MUC3B mucin (Fragment).
GN MUC3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pratt W.S., Crawley S., Hicks J., Ho J., Nash M., Kim Y.S., Gum J.R.,
RA Swallow D.M.;
RT "Multiple transcripts of MUC3: Evidence for two genes, MUC3A and
RT MUC3B.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ291390; CAC19572.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 1.
FT NON_TER 1
SQ SEQUENCE 901 AA; 96192 MW; F98FC817494ECD99 CRC64;

Query Match      81.5%; Score 44; DB 4; Length 901;
Best Local Similarity 70.0%; Pred. No. 8.2;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 549 CLPGFSGDRC 558
I: |:|||||
```

```

RESULT 11
O14651
ID O14651 PRELIMINARY; PRT; 957 AA.
AC O14651;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-2001 (TrEMBLrel. 17, Last annotation update)
DE Mucin (Intestinal mucin) (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476275; PubMed=9334251;
RA Gum J.R. Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA Vinall L.E., Robertson A.M., Swallow D.M., Kim Y.S.;
RT "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
RT terminus, and a novel upstream repetitive region.";
RL J. Biol. Chem. 272:26678-26686(1997).
DR EMBL; AF007194; AAC02272.1; -
DR EMBL; AF007196; AAB84383.1; -
DR EMBL; AF007195; AAB84382.1; -
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 957 AA; 9554 MW; E67722560D600A1F CRC64;

Query Match 81.5%; Score 44; DB 4; Length 957;
Best Local Similarity 70.0%; Pred. No. 8.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 865 CLPGFSGDRC 874

RESULT 12
Q9UKW9
ID Q9UKW9 PRELIMINARY; PRT; 1217 AA.
AC Q9UKW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Intestinal mucin 3 (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=INTESTINE;
RA Crawley S.C., Gum J.R. Jr., Hicks J.W., Pratt W.S., Aubert J.P.,
RA Swallow D.M., Kim Y.S.;
RT "Genomic organization and structure of the 3' region of human MUC3:
RT alternative splicing predicts membrane-bound and soluble forms of the
RT mucin.";
RL Biochem. Biophys. Res. Commun. 263:728-736(1999).
DR EMBL; AF113616; AAF13032.1; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.

```

```

DR PROSITE; PS50024; SEA; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 1217 AA; 129171 MW; 0783A4003C969664 CRC64;

Query Match 81.5%; Score 44; DB 4; Length 1217;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 865 CLPGFSGDRC 874

RESULT 13
Q28867
ID Q28867 PRELIMINARY; PRT; 89 AA.
AC Q28867;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95000251; PubMed=7916972;
RA Stewart F., Power C.A., Lennard S.N., Allen W.R., Amet L.,
RA Edwards R.M.;
RT "Identification of the horse epidermal growth factor (EGF) coding
RT sequence and its use in monitoring EGF gene expression in the
RT endometrium of the pregnant mare.";
RL J. Mol. Endocrinol. 12:341-350(1994).
DR EMBL; S73527; AAB32226.1; -
DR HSSP; P01132; IEGF.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001336; EGF_1.
DR Pfam; PF000008; EGF; 2.
DR PRINTS; PR000009; EGFTGF.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF-like; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 2.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 89 AA; 9879 MW; A7856F5E870B4A4B CRC64;

Query Match 79.6%; Score 43; DB 6; Length 89;
Best Local Similarity 60.0%; Pred. No. 1.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDRC 10
Db 74 CVGVGVGERC 83

RESULT 14
Q20043
ID Q20043 PRELIMINARY; PRT; 264 AA.
AC Q20043;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F35D2.3 protein.
GN F35D2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;

```

```

[1]
RN  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RX  MEDLINE=94150718; PubMed=7906398;
RA  Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA  Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA  Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA  Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA  Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA  Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA  Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA  Smaison J., Smith A., Sonhammer E., Staden R., Sulston J.,
RA  Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA  Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT  "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL  elegans.";
RL  Nature 368:32-38(1994).
RN  [2]
RN  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RC  Connell M.;
RT  "The sequence of C. elegans cosmid F35D2.";
RL  Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN  [3]
RN  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RC  Waterston R.;
RL  Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RL  EMBL: U28741; AAA68325.1;
DR  HSP: P00740; IEDM.
DR  InterPro: IPR000561; EGF-like.
DR  InterPro: IPR002049; Laminin_EGF.
DR  SMART: SM00180; EGF_Lam; 1.
DR  SMART: SM00001; EGF_Like; 1.
DR  PROSITE: PS00022; EGF_1; UNKNOWN_5.
DR  PROSITE: PS01186; EGF_2; 2.
DR  PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW  EGF-like domain; Glycoprotein.
SQ  SEQUENCE 264 AA; 28434 MW; C91B1CC7B4B463F4 CRC64;

Query Match          79.6%; Score 43; DB 5; Length 264;
Best Local Similarity 70.0%; Pred. No. 3.5;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy  1 CVIGXSGDRC 10
    | | | | |
Db  76 CPVGLSGDRC 85

RESULT 15
Q90XX0
ID  Q90XX0      PRELIMINARY;      PRT;      165 AA.
AC  Q90XX0;
DT  01-DEC-2001 (TReMBLrel. 19, Created)
DT  01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT  01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE  New antigen receptor (Fragment).
OS  Orectolobus maculatus (spotted wobbegong).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC  Elasmobranchii; Galeomorphi; Galeoidea; Orectolobiformes;
OC  Orectolobidae; Orectolobus.
OX  NCBI_TaxID=168098;
RN  [1]
RN  SEQUENCE FROM N.A.
RX  MEDLINE=21451048; PubMed=11566324;
RA  Nuttall S.D., Krishnan U.V., Hattarki M., De Gori R., Irving R.A.,
RA  Hudson P.J.;
RT  "Isolation of the new antigen receptor from wobbegong sharks, and use
RT  as a scaffold for the display of protein loop libraries.";
RL  Mol. Immunol. 38:313-326(2001).
DR  EMBL: AF336090; AAK97358.1;
DR  InterPro: IPR003006; Ig_MHC.
DR  InterPro: IPR000504; RNA_rec_mot.
```

```

DR  Pfam; PF00047; Ig; 1.
DR  PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW  Receptor.
FT  NON_TER.      165      165
SQ  SEQUENCE 165 AA; 18259 MW; D4A81143C6FEDEICD CRC64;

Query Match          77.8%; Score 42; DB 13; Length 165;
Best Local Similarity 77.8%; Pred. No. 3.4;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy  2 VICXSGDRC 10
    | | | | |
Db  105 MIGYSGDRC 113

Search completed: July 2, 2003, 07:28:37
Job time : 42.3333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 43.3333 Seconds
(without alignments)
30.750 Million cell updates/sec

Title: US-09-673-785D-9
Perfect score: 56
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred.. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	10	21	AAV52143
2	56	100.0	47	23	AAAG66047
3	56	100.0	53	5	AAAP40315
4	56	100.0	53	10	AAAP91658
5	56	100.0	53	11	AAAR08007
6	56	100.0	53	16	AAAR67275
7	56	100.0	53	18	AAAR34466
8	56	100.0	53	18	AAAR34467
9	56	100.0	53	19	AAAW50134
10	56	100.0	53	19	AAAW50139

11	56	100.0	53	19	AAAW50140
12	56	100.0	53	22	AAAB37612
13	56	100.0	53	23	AAAE15714
14	56	100.0	54	23	AAU76704
15	56	100.0	54	23	AAU76706
16	56	100.0	56	7	AAAP60791
17	56	100.0	117	19	AAAG68455
18	56	100.0	376	22	AAU02933
19	56	100.0	376	22	AAU02956
20	56	100.0	444	22	AAU02934
21	56	100.0	576	22	AAU02930
22	56	100.0	576	22	AAU02955
23	56	100.0	1073	22	AAU02953
24	56	100.0	1080	6	AAAP50296
25	51	91.1	2146	22	ABAB62317
26	49	87.5	655	23	AAAS2242
27	49	87.5	741	21	AAAB43485
28	49	87.5	2409	12	AAAR12609
29	47	83.9	14	18	AAAW34460
30	46	82.1	46	23	AAAG66045
31	46	82.1	48	15	AAAR57106
32	46	82.1	48	20	AAAY01790
33	46	82.1	71	9	AAAP81372
34	46	82.1	73	9	AAAP81367
35	46	82.1	298	23	AAU11636
36	46	82.1	754	18	AAAW27536
37	45	80.4	17	13	AAAR21704
38	45	80.4	17	18	AAAW11620
39	45	80.4	22	18	AAAW11621
40	45	80.4	22	18	AAAW11622
41	45	80.4	35	22	AAAG98651
42	45	80.4	47	23	AAAG66048
43	45	80.4	48	14	AAAR44820
44	45	80.4	48	21	AAAB36715
45	45	80.4	49	10	AAAP90465

ALIGNMENTS

RESULT 1

AAV52143

ID AAV52143 standard; peptide; 10 AA.

XX AAV52143;

XX 28-JAN-2000 (first entry)

XX Mouse EGF derived peptide for targeting laminin receptor.

DE Epidermal growth factor; EGF; laminin receptor; angiogenesis;
KW medicament; wound healing; retinopathy of immaturity; metastatic cancer;
KW candida infection; leishmania; trichomonas vaginalis.

XX Mus sp.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "Acetyl-Cys (S-ACM)"

FT Modified-site 10 /note= "Cys (S-ACM)-NH2"

XX WO9954356-A1.

XX 28-OCT-1999.

XX 21-APR-1999; 99WO-GB01211.

XX 22-APR-1998; 98GB-0008407.

XX (UYBE-) UNIV QUEENS BELFAST.

XX Nelson J, Walker B, McFerran N, Harriott P;

Epidermal growth f
Human EGF. Homo s
Epidermal growth f
Mouse epidermal gr
Modified mouse epi
Epidermal growth f
Recombinant human
Angiotensin conve
Angiotensin conve
Angiotensin conve
Angiotensin conve
Angiotensin conve
Mouse epidermal gr
Drosophila melanog
Human versican iso
Human cancer assoc
Versican. Homo sa
Chimeric epidermal
Mouse NRG-2 EGF-11
N terminal of rat
Rat epithelial gro
Rat epidermal grow
Rat epidermal grow
Human Neuroligin-2
Rat cerebellum der
EGF fragment. Syn
Target peptide fro
Target peptide fro
Epidermal Growth F
Human EGF motif se
Nicked EGF1-48. H
EGF-like domain of
Epidermal growth f

XX WPI; 2000-013229/01.
 XX New peptide derived from murine epidermal growth factor (mEGF) -
 XX Claim 4; Page 28; 35pp; English.
 PS This is a peptide derived from mouse epidermal growth factor (EGF)
 CC residues 33-42. This peptide is used in the invention to prepare a
 CC composition to target laminin receptors. EGF derived peptides inhibit
 CC blood vessel formation through their antagonism of the high affinity 67kD
 CC laminin receptor found on endothelial cells. The peptide is modified
 CC from the natural sequence to prevent protease attack. The peptide is used
 CC in the preparation of a medicament for binding to laminin receptors as an
 CC (ant)agonist. The medicament is also useful for healing endothelial cell
 CC wounds and treating angiogenic diseases, especially retinopathy of
 CC immaturity. Other diseases treated include metastatic cancer.
 CC Candida spp. infection, and parasitic infestations like leishmania and
 CC trichomonas vaginalis. The peptide are anti-angiogenic in human models.
 CC The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,
 CC and prevent tumour cell attachment to basement membranes.
 XX
 SQ Sequence 10 AA;
 Query Match 100.0%; Score 56; DB 21; Length 10;
 Best Local Similarity 90.0%; Pred. NO. 0.064;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 Db 1 CVIGYSGDRC 10
 |||||:|
 |||||:|
 RESULT 2
 AAG6047
 ID AAG66047 standard; peptide; 47 AA.
 AC AAG66047;
 XX
 XX 27-FEB-2002 (first entry)
 DT
 DE Mouse EGF motif sequence.
 DE
 KW ErbB-4; neuregulin-4; NRG-4; pro-NRG-4; neuroprotective; vulnery;
 KW cerebroprotective; vasotropic; antiparkinsonian; anticonvulsant;
 KW cytostatic; nootropic; EGF; NRG-2.
 XX
 OS Mus musculus.
 XX
 PN WO200181540-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-IL00371.
 XX
 PR 21-APR-2000; 2000US-0553769.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Harari D, Yarden Y;
 XX
 DR WPI; 2002-041398/05.
 XX
 XX Novel ErbB-4 ligand, referred as neuregulin (NRG)-4 and polynucleotide
 PT sequences encoding NRG-4, useful for upregulating or downregulating
 PT ErbB-4 receptor activity to treat Alzheimer's disease, stroke, gastric
 PT cancer -
 XX
 PS Disclosure; Fig 1c; 153pp; English.
 XX
 CC The invention relates to a novel ErbB-4 ligand, neuregulin-4 (NRG-4).
 CC NRG-4 binds to mammalian ErbB-4 receptor and can be expressed by standard
 CC recombinant methodology. Pharmaceutical compositions comprising NRG-4 are

CC useful for regulating an endogenous protein affecting ErbB-4 receptor
 CC activity in vivo. They are also useful for treating or preventing a
 CC disease condition or syndrome associated with dysregulation of an
 CC endogenous protein affecting ErbB-4 receptor activity, e.g., amyotrophic
 CC lateral sclerosis (Lou Gehrig's disease), Bell's palsy, spinal muscular
 CC atrophy, brain trauma, stroke, ischemia, Alzheimer's disease, Parkinson's
 CC disease, epilepsy, multiple sclerosis, Huntington's chorea, Down's
 CC syndrome, nerve deafness, neuropathy, muscular dystrophy, extramammary
 CC Paget's disease, gastric, pancreatic, prostate, breast and ovarian
 CC cancer, cervical carcinoma, endometrial adenocarcinoma, pancreatic D
 CC cells-somatostatinoma and Zollinger-Ellison syndrome. The agent comprised
 CC in the pharmaceutical composition includes a polypeptide (e.g., a soluble
 CC ligand binding domain of ErbB-4 i.e., IgB4; or a monoclonal, polyclonal,
 CC humanized, single chain antibody or an immunoreactive derivative of an
 CC antibody) capable of binding the endogenous protein affecting ErbB-4
 CC receptor activity. Traceable synthetic/recombinant NRG-4-tagged molecules
 CC can serve as a diagnostic tool in which cells binding NRG-4 can be
 CC measured. Sequences AAG66044-53 represent the EGF-like motifs of various
 CC growth factors.
 XX

SQ Sequence 47 AA;
 Query Match 100.0%; Score 56; DB 23; Length 47;
 Best Local Similarity 90.0%; Pred. NO. 0.25;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
 Db 33 CVIGYSGDRC 42
 |||||:|
 |||||:|

RESULT 3

AAP40315
 ID AAP40315 standard; protein; 53 AA.
 AC AAP40315;
 XX

XX 22-JUL-1992 (first entry)
 DT
 DE Epidermal growth factor tripentacontapeptide.
 XX
 KW EGF.
 XX
 OS Synthetic.
 XX
 PF Key Location/Qualifiers
 FT Disulfide-bond 6..20
 FT Disulfide-bond 14..31
 FT Disulfide-bond 33..42
 XX
 PN JP59027858-A.
 XX
 PD 14-FEB-1984.
 XX
 PF 05-AUG-1982; 82JP-0137128.
 XX
 PR 05-AUG-1982; 82JP-0137128.
 XX
 PA (NNSH) NIPPON SHINYAKU KK.
 XX
 DR WPI; 1984-072465/12.
 XX

XX Synthesis of epidermal growth factor polypeptide - by condensation
 PT of protected smaller peptide sequences, de:protection then oxidn. to
 PT cyclise.
 XX
 PS Claim1; Page 1; 8pp; Japanese.
 XX
 CC The amino acid sequence is that of an epidermal growth factor
 CC tripentacontapeptide which is synthesised by condensation of
 CC protected smaller peptide sequences. This method produces the
 CC peptide smoothly, with high purity and yield.

SQ Sequence 53 AA;

Query Match 100.0%; Score 56; DB 5; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.28;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
 DB 33 CVIGYSGDRC 42
 |||||:|

RESULT 4

AAP91658
 ID AAP91658 standard; protein; 53 AA.

XX

AC AAP91658;

DT 29-JUN-1990 (first entry)

XX Amino acid sequence for naturally occurring epidermal growth factor (EGF).
 DE (EGF).
 XX

KW Epidermal growth factor; angiogenesis; synthetic peptide.

XX WO8901489-A.

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

KW epithelial wounds; gastric acid secretion.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

SQ Sequence 53 AA;

Query Match 100.0%; Score 56; DB 11; Length 53;

Best Local Similarity 90.0%; Pred. No. 0.28;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

DB 33 CVIGYSGDRC 42

|||||:|

RESULT 6

AAR67275

ID AAR67275 standard; peptide; 53 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

KW Query Match 100.0%; Score 56; DB 10; Length 53;

XX Best Local Similarity 90.0%; Pred. No. 0.28;

XX Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 CVIGYSGDXC 10

XX DB 33 CVIGYSGDRC 42

XX |||||:|

XX RESULT 5

XX AAR08007

XX ID AAR08007 standard; protein; 53 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

KW Query Match 100.0%; Score 56; DB 10; Length 53;

XX Best Local Similarity 90.0%; Pred. No. 0.28;

XX Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 CVIGYSGDXC 10

XX DB 33 CVIGYSGDRC 42

XX |||||:|

XX RESULT 5

XX AAR08007

XX ID AAR08007 standard; protein; 53 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

KW Query Match 100.0%; Score 56; DB 10; Length 53;

XX Best Local Similarity 90.0%; Pred. No. 0.28;

XX Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 CVIGYSGDXC 10

XX DB 33 CVIGYSGDRC 42

XX |||||:|

XX RESULT 5

XX AAR08007

XX ID AAR08007 standard; protein; 53 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 13-NOV-1989; 89US-0434305.
XX (LIND/) LINDSTROM R L.
PA (SKEL/) SKELNIK D.
XX
XX
PI Lindstrom RL, Skelnik D;
XX
XX WPI; 1995-005865/01.

XX Viscoelastic soln. contg. hydroxypropyl methylcellulose, sodium
PT hyaluronate, chondroitin sulphate and growth factors - used as a
PT surgical soln. to promote wound healing, e.g. of corneal cells
XX
XX

PS Disclosure; Column 5; 8pp; English.

XX The sequences given in AAR67273-76 are cell growth factors which may
CC be used in the viscoelastic solution of the invention. The peptides
CC are derived from fibroblastic growth factor (FGF) and epidermal
CC growth factor (EGF). The viscoelastic solution also comprises a
CC buffered solution which is pref. a buffered balanced salt solution,
CC at least 1 of hydroxy- or carboxypropyl methylcellulose, a cellulose
CC gum, dextran or dextran sulphate, chondroitin sulphate, and sodium
CC hyaluronate. It has a pH of 6-8 and an osmolality of 200-400 mOsmol/L.
CC The growth factors are mitogenic in vitro for a wide range of tissues
CC and the viscoelastic soln. may be used as a surgical soln. which is
CC in direct contact with cells undergoing wound healing. It also provides
CC a cell protection and cell coating during surgery. The soln. provides
CC maintenance of tissue space, hydroxypropyl methylcellulose and
CC chondroitin sulphate lubricate the tissue, while sodium hyaluronate
CC provides tissue manipulation.

XX Sequence 53 AA;

Query Match 100.0%; Score 56; DB 16; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.28;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGYSGDNC 10
XXXXXXXXXX
Db 33 CVIGYSGDNC 42

RESULT 7
AAW34466
ID AAW34466 standard; Protein; 53 AA.

XX AAW34466;

XX 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein ABC.

XX Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

XX Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

XX Key Location/Qualifiers
FH Region 1..10
FT /note= "region A from mouse EGF"

FT Region 11..33
FT /note= "region B from human EGF"

FT Region 34..47
FT /note= "region C from mouse EGF"

FT Region 48..53
FT /note= "region D from human EGF"

XX AU9717759-A.

XX 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.

PR 23-APR-1996; 96JP-0123970.

XX (HGET) HIGETA SHOYU KK.

XX Tagami H, Tanaka A;

XX WPI; 1997-550187/51.

DR N-PSDB; AAT99941.

XX Chimeric epidermal growth factor proteins - and DNA molecules for
PT their recombinant production

XX Claim 5; Page 26; 40pp; English.

XX This is a chimeric epidermal growth factor (EGF) protein AC. This
CC chimeric EGF protein contains sequences derived from human and mouse
CC EGF sequences. The chimeric EGF proteins can be recombinantly produced
CC by culturing a microorganism of the genus Bacillus transformed by an
CC expression vector containing the encoding DNA molecules. The recombinant
CC chimeric EGF proteins exhibit a wide variety of physiological activities
CC similar to those of naturally occurring EGF.

XX Sequence 53 AA;

Query Match 100.0%; Score 56; DB 18; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.28;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGYSGDNC 10
XXXXXXXXXX
Db 33 CVIGYSGDNC 42

RESULT 8
AAW34467

ID AAW34467 standard; Protein; 53 AA.

XX AAW34467;

XX 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein ABC.

XX Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

XX Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

XX Key Location/Qualifiers
FH Region 1..10
FT /note= "region A from mouse EGF"

FT Region 11..33
FT /note= "region B from mouse EGF"

FT Region 34..47
FT /note= "region C from mouse EGF"

FT Region 48..53
FT /note= "region D from human EGF"

XX AU9717759-A.

XX 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.

XX 23-APR-1996; 96JP-0123970.

XX (HGET) HIGETA SHOYU KK.

XX Tagami H, Tanaka A;

XX WPI; 1997-550187/51.

DR N-PSDB; AAT99942.

XX

PT Chimeric epidermal growth factor proteins - and DNA molecules for
 XX their recombinant production
 PS Claim 6; Page 26; 40pp; English.
 XX This is a chimeric epidermal growth factor (EGF) protein ABC. This
 CC chimeric EGF protein contains sequences derived from human and mouse
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced
 CC by culturing a microorganism of the genus *Bacillus* transformed by an
 CC expression vector containing the encoding DNA molecules. The recombinant
 CC chimeric EGF proteins exhibit a wide variety of physiological activities
 CC similar to those of naturally occurring EGF.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 56; DB 18; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.28;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 Db 33 CVIGYSGDRC 42
 |||||:|
 RESULT 9
 AAW50134
 ID AAW50134 standard; protein; 53 AA.
 XX
 AC AAW50134;
 XX
 DT 08-JUL-1998 (first entry)
 XX
 DE Mouse epidermal growth factor.
 XX
 KW Mouse; epidermal growth factor; EGF; isolation; recombinant;
 KW ion exchange chromatography; *Bacillus brevis*.
 XX
 OS Mus sp.
 XX
 PN AU9728698-A.
 XX
 PD 05-FEB-1998.
 XX
 PF 17-JUL-1997; 97AU-0028698.
 XX
 PR 01-AUG-1996; 96JP-0218109.
 XX
 PA (HGET) HIGETA SHOYU KK.
 XX
 PI Miyauchi A, Nemoto A, Warren B;
 XX
 DR WPI; 1998-131057/13.
 XX
 PT Isolation of recombinant epidermal growth factor from whole broth -
 PT by fluidised-bed ion-exchange chromatography
 XX
 PS Claim 6; Page 26; 38pp; English.
 XX
 CC The present sequence represents mouse epidermal growth factor (EGF)
 CC from the present invention. The present invention describes a method for
 CC the isolation of recombinant epidermal growth factor (EGF) from whole
 CC broth. The method comprises: (a) passing a culture containing the
 CC protein upwards through a column containing a fluidised bed of ion
 CC exchanger, washing the ion exchanger to remove non-adsorbed material,
 CC allowing the ion exchanger to settle, and eluting adsorbed protein by
 CC passing an eluant downwards through the column; (b) filtering the eluate
 CC through a membrane with a molecular weight cutoff of 10000, and (c)
 CC concentrating the filtrate on a membrane with a molecular weight cutoff
 CC of 5000. The method is for isolating recombinant human, mouse, pig or
 CC rat EGF, or an EGF analogue with one of four defined amino acid sequences
 CC (given in the specification), from *Bacillus brevis* cultures. The EGF
 CC proteins can be recovered in high yield and high purity without the need
 CC to pretreat the culture broth to remove the cells, e.g. by costly

CC centrifugation or membrane filtration.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 56; DB 19; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.28;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 Db 33 CVIGYSGDRC 42
 |||||:|
 RESULT 10
 AAW50139
 ID AAW50139 standard; protein; 53 AA.
 XX
 AC AAW50139;
 XX
 DT 08-JUL-1998 (first entry)
 XX
 DE Epidermal growth factor analogue 3.
 XX
 KW Epidermal growth factor; EGF; isolation; recombinant;
 KW ion exchange chromatography; *Bacillus brevis*.
 XX
 OS Unidentified.
 XX
 PN AU9728698-A.
 XX
 PD 05-FEB-1998.
 XX
 PF 17-JUL-1997; 97AU-0028698.
 XX
 PR 01-AUG-1996; 96JP-0218109.
 XX
 PA (HGET) HIGETA SHOYU KK.
 XX
 PI Miyauchi A, Nemoto A, Warren B;
 XX
 DR WPI; 1998-131057/13.
 XX
 PT Isolation of recombinant epidermal growth factor from whole broth -
 PT by fluidised-bed ion-exchange chromatography
 XX
 PS Claim 6; Page 28; 38pp; English.
 XX
 CC The present sequence represents a protein which has epidermal growth
 CC factor (EGF) like activity, from the present invention. The present
 CC invention describes a method for the isolation of recombinant epidermal
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing
 CC a culture containing the protein upwards through a column containing a
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-
 CC adsorbed material, allowing the ion exchanger to settle, and eluting
 CC adsorbed protein by passing an eluant downwards through the column; (b)
 CC filtering the eluate through a membrane with a molecular weight cutoff
 CC of 10000, and (c) concentrating the filtrate on a membrane with a
 CC molecular weight cutoff of 5000. The method is for isolating recombinant
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four
 CC defined amino acid sequences (given in the specification), from *Bacillus*
 CC *brevis* cultures. The EGF proteins can be recovered in high yield and
 CC high purity without the need to pretreat the culture broth to remove the
 CC cells, e.g. by costly centrifugation or membrane filtration.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 56; DB 19; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.28;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 Db 33 CVIGYSGDRC 42
 |||||:|

```

RESULT 11
AAW50140
ID AAW50140 standard; protein; 53 AA.
XX
XX AAW50140;
AC
XX
XX 08-JUL-1998 (first entry)
XX
XX Epidermal growth factor analogue 4.
XX
XX Epidermal growth factor; EGF; isolation; recombinant;
KW ion exchange chromatography; Bacillus brevis.
XX
XX Unidentified.
OS
XX AU9728698-A.
XX
XX 05-FEB-1998.
PD
XX
XX 17-JUL-1997; 97AU-0028698.
PF
XX
XX 01-AUG-1996; 96JP-0218109.
PR
XX
XX (HGET ) HIGETA SHOYU KK.
PA
XX
XX Miyauchi A, Nemoto A, Warren B;
PI
XX
XX WPI; 1998-131057/13.
DR
XX
XX Isolation of recombinant epidermal growth factor from whole broth -
PT by fluidised-bed ion-exchange chromatography
PT
XX
XX Claim 6; Page 29; 38pp; English.
PS
XX
XX The present sequence represents a protein which has epidermal growth
CC factor (EGF) like activity, from the present invention. The present
CC invention describes a method for the isolation of recombinant epidermal
CC growth factor (EGF) from whole broth. The method comprises: (a) passing
CC a culture containing the protein upwards through a column containing a
CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-
CC adsorbed material, allowing the ion exchanger to settle, and eluting
CC adsorbed protein by passing an eluent downwards through the column; (b)
CC filtering the eluate through a membrane with a molecular weight cutoff
CC of 10000, and (c) concentrating the filtrate on a membrane with a
CC molecular weight cutoff of 5000. The method is for isolating recombinant
CC human, mouse, pig or rat EGF, or an EGF analogue with one of four
CC defined amino acid sequences (given in the specification), from Bacillus
CC brevis cultures. The EGF proteins can be recovered in high yield and
CC high purity without the need to pretreat the culture broth to remove the
CC cells, e.g. by costly centrifugation or membrane filtration.
XX
XX
XX Sequence 53 AA;
SQ
Query Match 100.0%; Score 56; DB 19; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.28;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10
DB 33 CVIGYSGDRC 42
|||||||:|

RESULT 13
AAE15714
ID AAE15714 standard; Protein; 53 AA.
XX
XX AAE15714;
AC
XX
XX 12-MAR-2002 (first entry)
DT
XX
XX Epidermal growth factor (EGF) used in exemplification of the invention.
DE
XX
XX Humanised form; monoclonal antibody alpha 340; gene therapy;
KW epidermal growth factor receptor; EGF; cancer; colorectal; lung; breast;
KW gastric; ovarian; immune response; cytostatic; cell growth; apoptosis;
KW inhibitor.
XX
XX Unidentified.
OS
XX WO200188138-A1.
PN
XX
XX 22-NOV-2001.
PD

Merozoite surface protein; protazoacide; vaccine; malaria; human; EGF;
epidermal growth factor.
XX
XX Homo sapiens.
XX
XX WO200063245-A2.
PN
XX
XX 26-OCT-2000.
PD
XX
XX 20-APR-2000; 2000WO-GB01558.
PF
XX
XX 20-APR-1999; 99GB-0009072.
PR
XX 13-MAY-1999; 99US-0311817.
PR
XX 25-MAY-1999; 99CA-2271451.
PR
XX
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX
XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiplbull C;
PI
XX
XX WPI; 2001-015762/02.
DR
XX
XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria -
PT
XX
XX Disclosure; Fig 1; 126pp; English.
PS
XX
XX The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of
CC the present invention are useful for immunising a mammal against malaria,
CC and can be used to treat malaria. The present sequence is human
CC epidermal growth factor (EGF). This sequence was used in a sequence
CC homology comparison with the wild-type MSP-1 protein from P. falciparum
CC (see AAB37608), which was used to generate the variants of the present
CC invention.
XX
XX
XX Sequence 53 AA;
SQ
Query Match 100.0%; Score 56; DB 22; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.28;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10
DB 33 CVIGYSGDRC 42
|||||||:|

RESULT 12
AAB37612
ID AAB37612 standard; protein; 53 AA.
XX
XX AAB37612;
AC
XX
XX 27-FEB-2001 (first entry)
DT
XX
XX Human EGF.
DE

```

XX PF 21-MAY-2001; 2001WO-GB02226.
 XX PR 19-MAY-2000; 2000GB-0011981.
 XX PR 24-AUG-2000; 2000GB-0020794.
 XX PA (SCAN-) SCANCELL LTD.
 XX PI Ellis JRM, Durrant LG;
 XX PI WPI; 2002-062384/08.
 XX DR
 XX PT New humanized form of mouse monoclonal antibody 340 which binds to
 XX PT epidermal growth factor receptor and inhibits binding of growth factor,
 XX PT useful for treating colorectal, lung, breast, gastric and ovarian
 XX PT cancer.
 XX PS Example 5; Fig 12; 53pp; English.
 XX CC The present invention relates to a humanised form of the antibody 340 (a
 CC mouse monoclonal antibody which binds to epidermal growth factor (EGF)
 CC receptor and inhibits binding of EGF), obtainable from the cell line
 CC deposited with the ECACC under accession number 97021428. The humanised
 CC form of the antibody 340 is useful in gene therapy, medicine and in the
 CC manufacture of a medicament for treatment or prophylaxis of cancer. The
 CC invention is useful for treating colorectal, lung, breast, gastric or
 CC ovarian cancers or also for preventing the recurrence of cancer after
 CC initial treatment or surgery. The invention is also useful for enhancing
 CC a protective immune response against cancer by optimised immunisation
 CC schedules. The humanised form of the antibody 340 has reduced
 CC immunogenicity but shows similar binding to cells expressing EGF
 CC receptor, as the original murine antibody and has increased ability to
 CC inhibit the growth of EGF receptor expressing cells. The invention is
 CC used as cell growth and apoptosis inhibitor. The present sequence
 CC is epidermal growth factor (EGF) which is used in the exemplification of
 CC the invention.
 XX SQ Sequence 53 AA;
 Query Match 100.0%; Score 56; DB 23; Length 53;
 Best Local Similarity 90.0%; Pred. No. 0.28;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 Db 33 CVIGYSGDRC 42
 RESULT 14
 AAU76704
 ID AAU76704 standard; Protein; 54 AA.
 AC AAU76704;
 XX 21-MAY-2002 (first entry)
 XX Mouse epidermal growth factor (EGF).
 XX DE
 XX KW Mouse; epidermal growth factor; receptor; EGF;
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;
 XX recombination.
 XX OS Mus sp.
 XX Key Location/Qualifiers
 FH Misc-difference 54
 FT /label= Unknown
 FT /note= "Encoded by TAA"
 XX WO200206469-A2.
 XX PD 24-JAN-2002.
 XX

PF 18-JUL-2001; 2001WO-US22640.
 XX PR 18-JUL-2000; 2000US-218921P.
 XX PR 18-JUL-2000; 2000US-219085P.
 XX PR 19-OCT-2000; 2000US-0691873.
 XX PR 19-OCT-2000; 2000US-0692732.
 XX PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.
 XX PI Coco WM, Ensell LP, Arensdorf JJ;
 XX PI WPI; 2002-179795/23.
 XX DR N-PSDB; ABK10313.
 XX PT Forming chimeric polynucleotide by contacting single-stranded (ss)
 XX PT scaffold fragment, and donor fragment populations to form hybridised
 XX PT complex having ss regions which are filled-in, and ligating adjacent
 XX PT fragments.
 XX PS Example 2; Fig 3; 65pp; English.
 XX CC The invention describes a method of forming (M1) chimeric
 CC polynucleotides (I). The method comprises contacting a population of
 CC single-stranded scaffold fragments (SF) with population of donor
 CC fragments (DF) to form a complex (II) comprising at least one SF
 CC hybridised to two DFs. (II) is treated so that single stranded regions of
 CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for
 CC a directed evolution process which involves forming a library of (I)
 CC that can be screened for a characteristic of interest. Subsequent rounds
 CC of directed evolution can produce chimeric polynucleotides with an
 CC improved characteristic of interest. The methods facilitate the
 CC generation of chimeric polynucleotides and do not require hybridising
 CC donor fragments to a target- or full-length template. Because the
 CC chimeraogenesis process does not rely upon a contiguous, full-length
 CC template, it is unnecessary to modify a template to facilitate its
 CC removal. This is the amino acid sequence of the mouse epidermal growth
 CC factor (EGF) used to demonstrate a method of in vitro recombination
 CC described in the method of the invention.
 XX SQ Sequence 54 AA;
 Query Match 100.0%; Score 56; DB 23; Length 54;
 Best Local Similarity 90.0%; Pred. No. 0.28;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 Db 33 CVIGYSGDRC 42
 RESULT 15
 AAU76706
 ID AAU76706 standard; Protein; 54 AA.
 AC AAU76706;
 XX 21-MAY-2002 (first entry)
 XX DE Modified mouse epidermal growth factor (EGF).
 XX KW Mouse; epidermal growth factor; receptor; EGF;
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;
 XX recombination.
 XX OS Mus sp.
 XX Key Location/Qualifiers
 FH Misc-difference 38
 FT /note= "Encoded by ACC"
 XX WO200206469-A2.
 XX

PD	24-JAN-2002.
XX	
PF	18-JUL-2001; 2001WO-US22640.
XX	
PR	18-JUL-2000; 2000US-218921P.
PR	18-JUL-2000; 2000US-219085P.
PR	19-OCT-2000; 2000US-0691873.
PR	19-OCT-2000; 2000US-0692732.
XX	
PA	(ENCH-) ENCHIRA BIOTECHNOLOGY CORP.
XX	
PI	Coco WM, EnCELL LP, Arensdorf JJ;
XX	
DR	WPI; 2002-179795/23.
DR	N-PSDB; ABK10315.
XX	
PT	Forming chimeric polynucleotide by contacting single-stranded (ss)
PT	scaffold fragment, and donor fragment populations to form hybridised
PT	complex having ss regions which are filled-in, and ligating adjacent
PT	fragments -
XX	
PS	Example 2; Fig 3; 65pp; English.
XX	
CC	The invention describes a method of forming (M1) chimeric
CC	polynucleotides (I). The method comprises contacting a population of
CC	single-stranded scaffold fragments (SF) with population of donor
CC	fragments (DF) to form a complex (II) comprising at least one SF
CC	hybridised to two DFs. (II) is treated so that single stranded regions of
CC	(II) are filled-in and adjacent fragments are ligated. (M1) is useful for
CC	a directed evolution process which involves forming a library of (I)
CC	that can be screened for a characteristic of interest. Subsequent rounds
CC	of directed evolution can produce chimeric polynucleotides with an
CC	improved characteristic of interest. The methods facilitate the
CC	generation of chimeric polynucleotides and do not require hybridising
CC	donor fragments to a target- or full-length template. Because the
CC	chimeragenesis process does not rely upon a contiguous, full-length
CC	template, it is unnecessary to modify a template to facilitate its
CC	removal. This is the amino acid sequence of a mouse epidermal growth
CC	factor (EGF) modified to make it as similar as possible to the human
CC	protein (AAU76705) and used to demonstrate a method of in vitro
CC	recombination described in the method of the invention.
XX	
SQ	Sequence 54 AA;
	Query Match 100.0%; Score 56; DB 23; Length 54;
	Best Local Similarity 90.0%; Pred. No. 0.28;
	Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	1 CVIGISGDCX 10
Db	33 CVIGISGDRC 42

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 10.6667 Seconds
(without alignments)
27.584 Million cell updates/sec

Title: US-09-673-785D-9
Perfect score: 56
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	12	1 US-08-597-545-7	Sequence 7, Appli
2	56	100.0	12	1 US-08-457-135-7	Sequence 7, Appli
3	56	100.0	53	1 US-08-284-923-2	Sequence 2, Appli
4	56	100.0	53	4 US-08-619-032B-2	Sequence 2, Appli
5	56	100.0	53	6 5332669-2	Patent No. 5332669
6	49	87.5	2409	6 5180808-2	Patent No. 5180808
7	48	85.7	154	3 US-09-191-647-10	Sequence 10, Appl
8	48	85.7	154	4 US-09-540-245A-10	Sequence 10, Appl
9	48	85.7	154	4 US-09-540-153-10	Sequence 10, Appl
10	46	82.1	48	6 5332669-1	Patent No. 5332669
11	46	82.1	754	2 US-08-525-864A-2	Sequence 2, Appli
12	45	80.4	17	6 5183805-1	Patent No. 5183805
13	45	80.4	17	6 5183805-2	Patent No. 5183805
14	45	80.4	25	6 5256643-8	Patent No. 5256643
15	45	80.4	37	2 US-08-039-364-14	Sequence 14, Appl
16	45	80.4	37	4 US-09-158-710-14	Sequence 14, Appl
17	45	80.4	42	1 US-08-168-091A-44	Sequence 44, Appl
18	45	80.4	44	1 US-08-278-089A-24	Sequence 24, Appl
19	45	80.4	46	3 US-08-899-437-12	Sequence 12, Appl
20	45	80.4	46	4 US-09-126-121-12	Sequence 12, Appl
21	45	80.4	47	3 US-08-753-007A-17	Sequence 17, Appl
22	45	80.4	47	4 US-09-398-496-17	Sequence 17, Appl
23	45	80.4	48	4 US-09-020-880-15	Sequence 15, Appl
24	45	80.4	48	4 US-08-915-096A-13	Sequence 13, Appl
25	45	80.4	48	4 US-09-101-544-15	Sequence 15, Appl
26	45	80.4	48	6 5434135-3	Patent No. 5434135
27	45	80.4	51	6 5177197-50	Patent No. 5177197

28	45	80.4	53	1 US-07-869-176-1	Sequence 1, Appli
29	45	80.4	53	1 US-08-284-923-1	Sequence 1, Appli
30	45	80.4	53	1 US-08-360-841-2	Sequence 2, Appli
31	45	80.4	53	2 US-08-861-000-1	Sequence 1, Appli
32	45	80.4	53	4 US-08-619-032B-1	Sequence 1, Appli
33	45	80.4	53	6 5332669-3	Patent No. 5332669
34	45	80.4	53	6 5434135-2	Patent No. 5434135
35	45	80.4	55	6 5218093-1	Patent No. 5218093
36	45	80.4	88	2 US-07-885-089B-13	Sequence 13, Appl
37	45	80.4	91	1 US-07-847-743B-15	Sequence 15, Appl
38	45	80.4	91	1 US-08-456-201-15	Sequence 15, Appl
39	45	80.4	91	2 US-08-330-161-13	Sequence 13, Appl
40	45	80.4	91	2 US-08-456-241-15	Sequence 15, Appl
41	45	80.4	91	2 US-08-440-401-13	Sequence 13, Appl
42	45	80.4	91	2 US-08-419-878B-13	Sequence 13, Appl
43	45	80.4	91	4 US-09-173-480-13	Sequence 13, Appl
44	45	80.4	91	5 PCT-US92-04295A-15	Sequence 15, Appl
45	45	80.4	93	1 US-08-343-401A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-597-545-7
; Sequence 7, Application US/08597545
; Patent No. 5580738
; GENERAL INFORMATION:
; APPLICANT: LABORDA, Jorge
; TITLE OF INVENTION: Delta-Like Gene Expressed In
; TITLE OF INVENTION: Neuroendocrine Tumors
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,545
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/989,537
; FILING DATE: 11-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/166 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-597-545-7

Query Match 100.0%; Score 56; DB 1; Length 12;
Best Local Similarity 90.0%; Pred. No. 0.011;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|||||||:|

Db 3 CVIGYSGDRC 12

RESULT 2
US-08-457-135-7
; Sequence 7, Application US/08457135
; Patent No. 5644031
; GENERAL INFORMATION:
; APPLICANT: LABORDA, Jorge
; TITLE OF INVENTION: Delta-Like Gene Expressed In
; TITLE OF INVENTION: Neuroendocrine Tumors
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,135
; FILING DATE: 01-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/989,537
; FILING DATE: 11-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/304/NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-457-135-7

Query Match 100.0%; Score 56; DB 1; Length 12;
Best Local Similarity 90.0%; Pred. No. 0.011;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
| | | | | | | | | |
Db 3 CVIGYSGDRC 12

RESULT 3
US-08-284-923-2
; Sequence 2, Application US/08284923
; Patent No. 5547935
; GENERAL INFORMATION:
; APPLICANT: Mullenbach, Guy T
; APPLICANT: Blaney, Jeffrey M
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: Muteins of Epidermal Growth Factor
; TITLE OF INVENTION: exhibiting enhanced binding at low ph
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,923
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/746,651
; FILING DATE: 16-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 231,001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-284-923-2

Query Match 100.0%; Score 56; DB 1; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.045;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
| | | | | | | | | |
Db 33 CVIGYSGDRC 42

RESULT 4
US-08-619-032B-2
; Sequence 2, Application US/08619032B
; Patent No. 6191106
; GENERAL INFORMATION:
; APPLICANT: Mullenbach, Guy T.
; APPLICANT: Blaney, Jeffrey M.
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: MUTEINS OF EPIDERMAL GROWTH FACTOR
; TITLE OF INVENTION: EXHIBITING ENHANCED BINDING AT LOW PH
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: Intellectual Property R-440, P.O. Box 8097
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,032B
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Guth, Joseph H.
; REGISTRATION NUMBER: 31,261
; REFERENCE/DOCKET NUMBER: 0231.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3888
; TELEFAX: (510) 655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-619-032B-2

Query Match 100.0%; Score 56; DB 4; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.045;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
||| ||||| :
Db 33 CVIGYSGDRC 42

RESULT 5

5332669-2
; Patent No. 5332669
; APPLICANT: DEUEL, THOMAS F.
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/356,739
; FILING DATE: 24-MAR-1989
; SEQ ID NO: 2:
; LENGTH: 53
5332669-2

Query Match 100.0%; Score 56; DB 6; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.045;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
||| ||||| :
Db 33 CVIGYSGDRC 42

RESULT 6

5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERISCAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERISCAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO: 2:
; LENGTH: 2409
5180808-2

Query Match 87.5%; Score 49; DB 6; Length 2409;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
||| ||||| :
Db 2128 CVPGYSGDQC 2137

RESULT 7

US-09-191-647-10
; Sequence 10, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; TITLE OF INVENTION: Tessler-Lavigne, Marc
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; PRIOR FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544

;
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-191-647-10

Query Match 85.7%; Score 48; DB 3; Length 154;
Best Local Similarity 60.0%; Pred. No. 1.9;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
||| ||||| :
Db 11 CMLGYTGDNC 20

RESULT 8

US-09-540-245A-10
; Sequence 10, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessler-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-540-245A-10

Query Match 85.7%; Score 48; DB 4; Length 154;
Best Local Similarity 60.0%; Pred. No. 1.9;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
||| ||||| :
Db 11 CMLGYTGDNC 20

RESULT 9

US-09-540-153-10
; Sequence 10, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessler-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-540-153-10

Query Match      85.7%; Score 48; DB 4; Length 154;
Best Local Similarity 60.0%; Pred. No. 1.9;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 11 CMLGYTGDC 20

RESULT 10
5332669-1
; Patent No. 5332669
; APPLICANT: DEUEL, THOMAS F.
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/356,739
; FILING DATE: 24-MAR-1989
; SEQ ID NO:1:
; LENGTH: 48
5332669-1

Query Match      82.1%; Score 46; DB 6; Length 48;
Best Local Similarity 70.0%; Pred. No. 1.3;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 33 CVIGYIGERC 42

RESULT 11
US-08-525-864A-2
; Sequence 2, Application US/08525864A
; Patent No. 5912326
; GENERAL INFORMATION:
; APPLICANT: Chand, Han
; TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses
; TITLE OF INVENTION: Related thereto
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,864A
; FILING DATE: 8-SEP-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kara, Catherine J.
; REGISTRATION NUMBER: 41,106
; REFERENCE/DOCKET NUMBER: HUI-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-525-864A-2

Query Match      82.1%; Score 46; DB 2; Length 754;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 280 CPVGYTGDC 289

RESULT 12
5183805-1
; Patent No. 5183805
; APPLICANT: LEE, JIN S.; BLICK, MARK
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/567,407
; FILING DATE: 13-AUG-1990
; SEQ ID NO:1:
; LENGTH: 17
5183805-1

Query Match      80.4%; Score 45; DB 6; Length 17;
Best Local Similarity 60.0%; Pred. No. 0.66;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 2 CVVGYIGERC 11

RESULT 13
5183805-2
; Patent No. 5183805
; APPLICANT: LEE, JIN S.; BLICK, MARK
; TITLE OF INVENTION: BIOACTIVE EGF PEPTIDES FOR
; PROMOTION OF TISSUE REGENERATION AND CANCER THERAPY
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/567,407
; FILING DATE: 13-AUG-1990
; SEQ ID NO:2:
; LENGTH: 17
5183805-2

Query Match      80.4%; Score 45; DB 6; Length 17;
Best Local Similarity 60.0%; Pred. No. 0.66;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 2 CVVGYIGERC 11

RESULT 14
5256643-8
; Patent No. 5256643
; APPLICANT: Persico, Maria G.; Salomon, David S.
; TITLE OF INVENTION: HUMAN CRYPTO PROTEIN
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/530,165
; FILING DATE: 29-MAY-1990
; SEQ ID NO:8:
; LENGTH: 25
5256643-8

Query Match      80.4%; Score 45; DB 6; Length 25;
```

Best Local Similarity 60.0%; Pred. No. 0.96;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 13 CVVGYIGERC 22

RESULT 15
US-08-039-364-14
; Sequence 14, Application US/08039364
; Patent No. 5811393
; GENERAL INFORMATION:
; APPLICANT: Klagsbrun, Michael
; APPLICANT: Abraham, Judith A.
; APPLICANT: Higashiyama, Shigeki
; APPLICANT: Besner, Gail F.
; TITLE OF INVENTION: HEPARIN BINDING MITOGEN WITH
; TITLE OF INVENTION: HOMOLOGY TO EPIDERMAL GROWTH FACTOR
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55Sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/039,364
; FILING DATE: 15 JUN 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/598,082
; FILING DATE: 16 OCT 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05162/002002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-039-364-14

Query Match 80.4%; Score 45; DB 2; Length 37;
Best Local Similarity 60.0%; Pred. No. 1.4;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 28 CVVGYIGERC 37

Search completed: July 2, 2003, 07:25:43
Job time : 11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:28:49 ; Search time 103 Seconds
(without alignments)
11.164 Million cell updates/sec

Title: US-09-673-785D-9
Perfect score: 56
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues
Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	56	100.0	53	9 US-10-138-158-19	Sequence 19, Appl
2	56	100.0	53	9 US-10-150-648B-33	Sequence 33, Appl
3	56	100.0	145	9 US-10-150-648B-30	Sequence 30, Appl
4	49	87.5	741	10 US-09-925-301-930	Sequence 930, Appl
5	46	82.1	112	9 US-10-211-994-20	Sequence 20, Appl
6	46	82.1	298	10 US-09-864-675-4	Sequence 4, Appl1
7	45	80.4	46	9 US-10-136-573A-12	Sequence 12, Appl
8	45	80.4	46	9 US-09-877-665-12	Sequence 12, Appl
9	45	80.4	46	9 US-10-215-862-12	Sequence 12, Appl
10	45	80.4	46	10 US-09-817-647-12	Sequence 12, Appl
11	45	80.4	47	12 US-10-096-241-17	Sequence 17, Appl
12	45	80.4	48	9 US-10-201-945-13	Sequence 13, Appl
13	45	80.4	53	9 US-09-903-327A-9	Sequence 9, Appl1
14	45	80.4	53	9 US-10-150-648B-35	Sequence 35, Appl
15	45	80.4	53	9 US-10-211-994-4	Sequence 4, Appl1
16	45	80.4	53	10 US-09-848-664-31	Sequence 31, Appl
17	45	80.4	58	10 US-09-934-706-3	Sequence 3, Appl1
18	45	80.4	91	9 US-10-022-609-13	Sequence 13, Appl
19	45	80.4	111	9 US-10-211-994-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-10-138-158-19
; Sequence 19, Application US/10138158
; Publication No. US20030036509A1
; GENERAL INFORMATION:
; APPLICANT: STEM CELL PHARMACEUTICALS, INC.
; APPLICANT: TWARDZIK, Daniel R.
; APPLICANT: PERNET, Andre
; APPLICANT: FELKER, Thomas S.
; APPLICANT: PASKELL, Stefan
; APPLICANT: RENO, John M.
; TITLE OF INVENTION: TGF-alpha POLYPEPTIDES, FUNCTIONAL FRAGMENTS AND METHODS OF
; FILE REFERENCE: STEM110-6
; CURRENT APPLICATION NUMBER: US/10/138.158
; CURRENT FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: US 09/641,587
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/559,248
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/459,813
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 09/492,935
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 09/378,567
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-158-19

Query Match 100.0%; Score 56; DB 9; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.064;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
| | | | | | | | | |
Db 33 CVIGYSGDXC 42

Sequence 29, Appl
Sequence 5, Appl
Sequence 7, Appl
Sequence 25, Appl
Sequence 64, Appl
Sequence 32, Appl
Sequence 5, Appl
Sequence 13, Appl
Sequence 6, Appl
Sequence 4, Appl
Sequence 7, Appl
Sequence 5, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 6253, Ap
Sequence 269, App
Sequence 31, Appl
Sequence 34, Appl
Sequence 266, App
Sequence 273, App
Sequence 272, App
Sequence 6, Appl
Sequence 8, Appl
Sequence 2, Appl
Sequence 171, App
Sequence 342, App

RESULT 2
US-10-150-648B-33
; Sequence 33, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725
; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 977 TO 1029
US-10-150-648B-33

Query Match 100.0%; Score 56; DB 9; Length 53;
Best Local Similarity 90.0%; Pred. No. 0.064; Indels 0; Gaps 0;
Matches 9; Conservative 1; Mismatches 0;
QY 1 CVIGYSGDXC 10
|||||||
DB 33 CVIGYSGDRC 42
RESULT 3
US-10-150-648B-30
; Sequence 30, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; TITLE: 128,000-molecular weight protein precursor
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725

DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 919 TO 1063
US-10-150-648B-30
Query Match 100.0%; Score 56; DB 9; Length 145;
Best Local Similarity 90.0%; Pred. No. 0.15; Indels 0; Gaps 0;
Matches 9; Conservative 1; Mismatches 0;
QY 1 CVIGYSGDXC 10
|||||||
DB 91 CVIGYSGDRC 100
RESULT 4
US-09-925-301-930
; Sequence 930, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 930
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (282)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-930
Query Match 87.5%; Score 49; DB 10; Length 741;
Best Local Similarity 80.0%; Pred. No. 7.2; Indels 0; Gaps 0;
Matches 8; Conservative 1; Mismatches 1;
QY 1 CVIGYSGDXC 10
|||||||
DB 460 CVPYSGDQC 469
RESULT 5
US-10-211-994-20
; Sequence 20, Application US/102111994
; Publication No. US20030082201A1
; GENERAL INFORMATION:
; APPLICANT: Rao, M.R.S.
; APPLICANT: Sengupta, Paromita
; APPLICANT: Prasad, Sudhanand
; APPLICANT: Burman, Anand C.
; APPLICANT: Mukherjee, Rama
; APPLICANT: Thomas, Becky
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER
; FILE REFERENCE: U014152-1
; CURRENT APPLICATION NUMBER: US/10/211,994
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/309,975
; PRIOR FILING DATE: 2001-08-03

NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide and peptide sequence of multivalent vaccine
US-10-211-994-20

Query Match 82.1%; Score 46; DB 9; Length 112;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 92 CVVGYIGERC 101

RESULT 6
US-09-864-675-4
; Sequence 4, Application US/09864675
; Patent No. US20020081286A1
; GENERAL INFORMATION:
; APPLICANT: Marchionni, Mark
; TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,
; POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS
; FILE REFERENCE: 04585/049002
; CURRENT APPLICATION NUMBER: US/09/864,675
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/206,495
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-675-4

Query Match 82.1%; Score 46; DB 10; Length 298;
Best Local Similarity 60.0%; Pred. No. 9.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 280 CPVGYTGDCR 289

RESULT 7
US-10-136-573A-12
; Sequence 12, Application US/10136573A
; Patent No. US20020161200A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; Uses Therefor
; FILE REFERENCE: P1084RIC2
; CURRENT APPLICATION NUMBER: US/10/136,573A
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: US 09/480,977
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 08/899,437
; PRIOR FILING DATE: 1997-07-24
; PRIOR APPLICATION NUMBER: US 60/052,019
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 12
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-136-573A-12
Query Match 80.4%; Score 45; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 2.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 32 CVVGYIGERC 41

RESULT 8
US-09-877-665-12
; Sequence 12, Application US/09877665
; Patent No. US20020164680A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/877,665
FILING DATE: 08-Jun-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/109,206
FILING DATE: 30-Jun-1998
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Delidre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
FEATURE:
NAME/KEY: hEGF.egf
LOCATION: 1-46
IDENTIFICATION METHOD:
OTHER INFORMATION:
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-877-665-12

Query Match 80.4%; Score 45; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 2.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
Db 32 CVVGYIGERC 41

RESULT 9
US-10-215-862-12
; Sequence 12, Application US/10215862
; Publication No. US20030036166A1
; GENERAL INFORMATION:

```

: APPLICANT: Godowski, Paul J.
: APPLICANT: Mark, Melanie Rose
: APPLICANT: Zhang, Dong Xiao
: TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
: TITLE OF INVENTION: Uses Therefor
: FILE REFERENCE: P1084R1D2C1
: CURRENT APPLICATION NUMBER: US/10/215,862
: CURRENT FILING DATE: 2002-09-24
: PRIOR APPLICATION NUMBER: US 09/126,663
: PRIOR FILING DATE: 1998-07-30
: PRIOR APPLICATION NUMBER: US 08/899,437
: PRIOR FILING DATE: 1997-07-24
: PRIOR APPLICATION NUMBER: US 60/052,019
: PRIOR FILING DATE: 1997-07-09
: NUMBER OF SEQ ID NOS: 23
: SEQ ID NO 12
: LENGTH: 46
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-215-862-12

Query Match      80.4%; Score 45; DB 9; Length 46;
Best Local Similarity 60.0%; Pred. No. 2.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps

QY      1  CVIGSGDXXC 10
        ||||| |::|
Db      32  CVVGIGERC 41

RESULT 10
US-09-817-647-12
: Sequence 12, Application US/09817647
: Patent No. US20020082229A1
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
: TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
: TITLE OF INVENTION: Ligands and Uses Therefor
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 1 DNA Way
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WinPatIn (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/817,647
: FILING DATE: 26-Mar-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/107,979
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Conley, Delirdre L.
: REGISTRATION NUMBER: 36,487
: REFERENCE/DOCKET NUMBER: P1084R1-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650/225-2066
: TELEFAX: 650/952-9881
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 46 amino acids
: TYPE: Amino Acid
: TOPOLOGY: Linear
: FEATURE:
: NAME/KEY: hEGF.egf
: LOCATION: 1-46

```


Publication No. US20020188110A1
GENERAL INFORMATION:
APPLICANT: Meissner, Paul S.
Fuldner, Rebecca A.
Adams, Mark D.
TITLE OF INVENTION: Transforming Growth Factor Alpha HI
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/201,945
FILING DATE: 25-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/471,377
FILING DATE: 23-Dec-1999
APPLICATION NUMBER: 08/915,096
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/208,008
FILING DATE: 08-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF110D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEX: 301-309-8439
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-201-945-13

Query Match 80.4%; Score 45; DB 9; Length 48;
Best Local Similarity 60.0%; Pred. No. 2.8;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYGDXC 10
Db 28 CVVGYIGERC 37

RESULT 13
US-09-903-327A-9
Sequence 9, Application US/09903327A
Patent No. US20020164333A1
GENERAL INFORMATION:
APPLICANT: Nemerow, Glen R.
Li, Erquang
TITLE OF INVENTION: BIFUNCTIONAL MOLECULES AND VECTORS COMPLEXED THEREWITH FOR TARGET
TITLE OF INVENTION: GENE
TITLE OF INVENTION: DELIVERY
FILE REFERENCE: 22908-1228
CURRENT APPLICATION NUMBER: US/09/903,327A
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 09/613,017
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9
LENGTH: 53
TYPE: PRT
ORGANISM: Human
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (0)...(0)
OTHER INFORMATION: Epidermal Growth Factor (EGF, mature peptide)
US-09-903-327A-9

Query Match 80.4%; Score 45; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. No. 3.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYGDXC 10
Db 33 CVVGYIGERC 42

RESULT 14
US-10-150-648B-35
Sequence 35, Application US/10150648B
Publication No. US20030059802A1
GENERAL INFORMATION:
APPLICANT: Bilodeau-Goesseels, Sylvie
APPLICANT: John, Sushil J.
APPLICANT: Selinger, Leonard B.
APPLICANT: Benkel, Bernhard F.
TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
factor
FILE REFERENCE: 60-01
CURRENT APPLICATION NUMBER: US/10/150,648B
CURRENT FILING DATE: 2002-05-17
PRIOR APPLICATION NUMBER: 60/292,136
PRIOR FILING DATE: 2001-05-18
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35
LENGTH: 53
TYPE: PRT
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
AUTHORS: Bell, G.I., Fong, N.M., Stempfen, M.M., Wormsted, M.A.,
AUTHORS: Caput, D., Ku, L.L., Urdea, M.S., Rall, L.B. and Sanchez-
AUTHORS: Pescador, R.
TITLE: Human epidermal growth factor precursor: cDNA sequence,
TITLE: expression in vitro and gene organization.
JOURNAL: Nucleic Acids Research
VOLUME: 14
ISSUE: 21
PAGES: 8427-8446
DATE: 1986
DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 X04571
DATABASE ENTRY DATE: 1993-04-21
RELEVANT RESIDUES: Relevant residues FROM 970 TO 1022
US-10-150-648B-35

Query Match 80.4%; Score 45; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. No. 3.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYGDXC 10
Db 33 CVVGYIGERC 42

RESULT 15
US-10-211-994-4
Sequence 4, Application US/10211994
Publication No. US20030082201A1
GENERAL INFORMATION:
APPLICANT: Rao, M.R.S.
APPLICANT: Sengupta, Paromita

```
; APPLICANT: Prasad, Sudhanand
; APPLICANT: Burman, Anand C.
; APPLICANT: Mukherjee, Rama
; APPLICANT: Thomas, Becky
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER
; FILE REFERENCE: U014152-1
; CURRENT APPLICATION NUMBER: US/10/211,994
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/309,975
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Epidermal growth factor
US-10-211-994-4

Query Match      80.4%; Score 45; DB 9; Length 53;
Best Local Similarity 60.0%; Pred. No. 3.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGYSGDXC 10
        ||:|||:|||
Db      33 CVVGIGERC 42

Search completed: July 2, 2003, 07:57:18
Job time : 103 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 12 Seconds
(without alignments)
80.112 Million cell updates/sec

Title: US-09-673-785D-9

Perfect score: 56

Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1217	1	EGMSG
2	51	91.1	2139	2	A35672
3	51	91.1	4006	2	T09070
4	49	87.5	102	2	B58885
5	49	87.5	372	2	T23559
6	49	87.5	862	2	S43922
7	49	87.5	1643	2	T14274
8	49	87.5	2409	1	A60979
9	49	87.5	3381	2	T42389
10	46	82.1	57	2	PC4415
11	46	82.1	1133	1	EGRT
12	46	82.1	1531	2	T42218
13	45	80.4	832	2	A31246
14	45	80.4	833	2	S19087
15	45	80.4	880	2	S00670
16	45	80.4	883	2	S43126
17	45	80.4	1207	1	EGHU
18	45	80.4	1722	2	E89753
19	45	80.4	2397	1	A55535
20	44	78.6	264	2	T16271
21	44	78.6	1064	2	A40136
22	44	78.6	1203	2	A43175
23	44	78.6	2352	2	T30201
24	44	78.6	2471	2	A49128
25	44	78.6	2476	2	T34022
26	44	78.6	3566	1	A40701
27	44	78.6	4135	2	T42629
28	43	76.8	373	2	T23300
29	43	76.8	2321	2	S78549

30 43 76.8 3672 2 T23433
31 43 76.8 3704 2 T37316
32 42 75.0 53 2 S17294
33 42 75.0 57 2 B69300
34 42 75.0 330 2 T46256
35 42 75.0 473 2 A56175
36 42 75.0 548 2 T25424
37 42 75.0 883 2 S57653
38 42 75.0 912 2 A54423
39 42 75.0 1408 2 S16148
40 42 75.0 1523 2 T13953
41 42 75.0 2318 2 S45306
42 42 75.0 2531 2 T31070
43 41 73.2 597 2 S71352
44 41 73.2 616 2 T29234
45 41 73.2 1049 1 S19421

ALIGNMENTS

RESULT 1

EGMSG

epidermal growth factor precursor - mouse

N;Alternate names: urogastrone precursor

C;Species: Mus musculus (house mouse)

C;Date: 30-Nov-1980 #sequence.Revision 11-Aug-1983 #text_change 19-Jan-2001

C;Accession: A94272; A93304; A92118; A01387

R;Scott, J.; Urdea, M.; Quiroga, M.; Sanchez-Pescador, R.; Fong, N.; Selby, M.; Rutt

Science 221, 236-240, 1983

A;Title: Structure of a mouse submaxillary messenger RNA encoding epidermal growth f

A;Reference number: A94272; MUID:83223630; PMID:6602382

A;Accession: A94272

A;Molecule type: mRNA

A;Residues: 1-1217 <SCO>

A;Cross-references: GB:J00380; NID:g192993; PIDN:AAA37539.1; PID:g309210

R;Gray, A.; Dull, T.; Ullrich, A.

Nature 303, 722-725, 1983

A;Title: Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-mole

A;Reference number: A93304; MUID:83219309; PMID:6304537

A;Accession: A93304

A;Molecule type: mRNA

A;Residues: 1-789, 'Y', 791-1047, 'S', 1049-1168 <GRA>

A;Cross-references: GB:J00380

A;Note: the sequence shown by these authors differs from residues 1134-1168 due to a

uence of Scott et al.)

R;Savage Jr., C.R.; Inagami, T.; Cohen, S.

J. Biol. Chem. 247, 7612-7621, 1972

A;Title: The primary structure of epidermal growth factor.

A;Reference number: A92118; MUID:73048516; PMID:4636327

A;Accession: A92118

A;Molecule type: protein

A;Residues: 977-1029 <SAV>

A;Note: residues 1024-1029 are not required for full biological activity in vivo

R;Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A;Title: Epidermal growth factor. Location of disulfide bonds.

A;Reference number: A92144; MUID:74025498; PMID:4750422

A;Contents: annotation; disulfide bonds

C;Comment: Epidermal growth factor (EGF) stimulates the proliferation and differenti

gastrointestinal cell proliferation.

C;Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in f

C;Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound pr

C;Comment: The active growth factor from this submaxillary gland protein stimulates

C;Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-co

C;Keywords: duplication; growth factor; tandem repeat; transmembrane protein

F;1-28/Domain: signal sequence #status predicted <SIG>

F;29-1217/Product: epidermal growth factor proprotein, membrane-bound form #status p

F;29-1038/Domain: extracellular #status predicted <EXT>

F;50-485/Region: EGF precursor long repeat

F;53-92/Domain: LDL receptor YWTD-containing repeat homology <YW01>

F;93-134/Domain: LDL receptor YWTD-containing repeat homology <YW02>

F:135-176/Domain: LDL receptor WYTD-containing repeat homology <YW03>
 F:177-217/Domain: LDL receptor WYTD-containing repeat homology <YW04>
 F:218-262/Domain: LDL receptor WYTD-containing repeat homology <YW05>
 F:263-307/Domain: LDL receptor WYTD-containing repeat homology <YW06>
 F:324-360/Domain: EGF homology #status atypical <EG1>
 F:366-401/Domain: EGF homology <EG2>
 F:407-442/Domain: EGF homology <EG3>
 F:445-482/Domain: EGF homology <EG4>
 F:486-529/Region: EGF precursor long repeat
 F:489-529/Domain: LDL receptor WYTD-containing repeat homology <YW07>
 F:530-572/Domain: LDL receptor WYTD-containing repeat homology <YW08>
 F:573-615/Domain: LDL receptor WYTD-containing repeat homology <YW09>
 F:616-659/Domain: LDL receptor WYTD-containing repeat homology <YW10>
 F:660-700/Domain: LDL receptor WYTD-containing repeat homology <YW11>
 F:701-743/Domain: LDL receptor WYTD-containing repeat homology <YW12>
 F:751-786/Domain: EGF homology <EG5>
 F:842-875/Domain: EGF homology <EG6>
 F:881-917/Domain: EGF homology <EG7>
 F:923-958/Domain: EGF homology <EG8>
 F:977-1029/Product: epidermal growth factor #status experimental <EGF>
 F:1039-1063/Domain: EGF homology <EG9>
 F:1039-1063/Domain: transmembrane #status predicted <TMW>
 F:1064-1217/Domain: intracellular #status predicted <INT>
 F:347-360,366-377,373-386,388-401,407-418,414-427,429-442,445-457,453-467,469-482,751-761
 tus predicted
 F:982-996,990-1007,1009-1018/Disulfide bonds: #status experimental

Query Match 100.0%; Score 56; DB 1; Length 1217;
 Best Local Similarity 90.0%; Pred. No. 0.12;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10
 | : ||||| : |
 Db 1009 CVIGYSGDRC 1018

RESULT 2

A35672
 crumbs protein - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 11-Jan-2000
 C:Accession: A35672
 R:Tapas, U.; Theres, C.; Knust, E.
 Cell 61, 787-799, 1990
 A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila
 A:Reference number: A35672; MUID:90263104; PMID:2344615
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2139 <TEP>
 A:Cross-references: GB:M33753
 A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue
 C:Genetics:

A:Gene: FlyBase:crb
 A:Cross-references: FlyBase:FBgn0000368
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: transmembrane protein
 F:352-385/Domain: EGF homology <EGX1>
 F:392-424/Domain: EGF homology <EGF1>
 F:691-722/Domain: EGF homology <EGF>
 F:767-799/Domain: EGF homology <EGF3>
 F:1878-1914/Domain: EGF homology <EGX2>

Query Match 91.1%; Score 51; DB 2; Length 2139;
 Best Local Similarity 70.0%; Pred. No. 1.5;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10
 | : ||||| : |
 Db 571 CAVGYSGDRC 580

RESULT 3

T09070
 probable tenascin X - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jan-2000
 C:Accession: T09070
 R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.;
 submitted to the EMBL Data Library, October 1997
 A:Description: Sequence of the mouse major histocompatibility locus class III region
 A:Reference number: Z16543
 A:Accession: T09070
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4006 <ROW>
 A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958
 C:Genetics:
 A:Gene: TNX
 A:Map position: 17
 A:Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1;
 019/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3730/3
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin
 C:Keywords: extracellular matrix
 F:422-448/Domain: EGF homology <EGF>
 F:422-448/Domain: EGF homology <EGF>
 F:826-906/Domain: fibronectin type III repeat homology <3FR>
 F:3789-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 91.1%; Score 51; DB 2; Length 4006;
 Best Local Similarity 70.0%; Pred. No. 2.7;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10
 | : ||||| : |
 Db 532 CAVGYSGDDC 541

RESULT 4

B55885
 chondroitin sulfate proteoglycan MV3 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 16-Jul-1999
 C:Accession: B55885
 R:Zako, M.; Shinomura, T.; Ujita, M.; Ito, K.; Kimata, K.
 J. Biol. Chem. 270, 3914-3918, 1995
 A:Title: Expression of PG-M(V3), an alternatively spliced form of PG-M without a canon
 A:Reference number: A55885; MUID:95181355; PMID:7876137
 A:Accession: B55885
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-102 <ZAK>
 A:Cross-references: GB:S75879; GB:D32039
 C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;
 F:24-55/Domain: EGF homology <EG1>
 F:62-93/Domain: EGF homology <EG2>

Query Match 87.5%; Score 49; DB 2; Length 102;
 Best Local Similarity 80.0%; Pred. No. 0.18;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDXC 10
 | : ||||| : |
 Db 46 CVPGYSGDQC 55

RESULT 5

T29359
 hypothetical protein R05G6.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29359
 R:Murray, J.; Le, T.T.
 submitted to the EMBL Data Library, May 1996
 A:Description: The sequence of C. elegans cosmid R05G6.
 A:Reference number: Z20612
 A:Accession: T29359

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-372 <EMBL>

A:Cross-references: EMBL:U58746; PIDN:AAB00626.1; GSPDB:GN00022; CESP:R05G6.9

A:Experimental source: strain Bristol N2; clone R05G6

C:Genetics:

A:Gene: CESP:R05G6.9

A:Map position: 4

A:Introns: 80/1; 161/1; 245/1; 286/1

Query Match 87.5%; Score 49; DB 2; Length 372;

Best Local Similarity 80.0%; Pred. No. 0.61;

Matches 7; Conservative 2; Mismatches 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|| |||||:|

Db 233 CVLGYSGDKC 242

RESULT 6

S43922

versican - pig-tailed macaque (fragments)

N:Alternate names: chondroitin sulfate proteoglycan

C:Species: Macaca nemestrina (pig-tailed macaque)

C>Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 16-Jul-1999

A:Accession: S43922

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43922

A:Molecule type: mRNA

A:Residues: 1-233;234-525;526-862 <YAO>

A:Cross-references: EMBL:572413

A>Note: 507-Ser was also found

A>Note: the authors translated the codon GCC for residue 50 as Val, AAG for residue 422

669 as Asn

C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; E

F:1-37/Domain: link protein repeat homology (fragment) <LNK1>

F:58-139/Domain: link protein repeat homology <LNK1>

F:722-753/Domain: EGF homology <EG1>

F:760-791/Domain: EGF homology <EG2>

Query Match

Best Local Similarity 87.5%; Score 49; DB 2; Length 862;

Matches 8; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|| |||||:|

Db 744 CVPYSGDQC 753

RESULT 7

T14274

versican precursor, splice form V2 - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000

A:Accession: T14274

R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.

J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine brain

A:Reference number: 217954; MUID:98288320; PMID:9624174

A:Accession: T14274

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1643 <SCH>

A:Cross-references: EMBL:AF060458; NID:g32533303; PID:g32533304; PIDN:AAC24360.1

A:Experimental source: brain

C:Keywords: glycoprotein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-1643/Product: versican, splice form V2 #status predicted <MAT>

F:57-331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent) #st

Query Match

Best Local Similarity 87.5%; Score 49; DB 2; Length 1643;

Matches 8; Conservative 1; Mismatches 0; Gaps 0;

QY 1 CVIGYSGDXC 10

|| |||||:|

Db 1362 CVPYSGDRC 1371

RESULT 8

A60979

versican precursor - human

N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteogly

C:Species: Homo sapiens (man)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179

R:Zimmermann, D.R.; Ruoslahti, E.

EMBO J. 8, 2975-2981, 1989

A:Title: Multiple domains of the large fibroblast proteoglycan, versican.

A:Reference number: S06014; MUID:90059882; PMID:2583089

A:Accession: S06014

A:Molecule type: mRNA

A:Residues: 1-2409 <ZIM>

A:Cross-references: GB:X15998; NID:g37662; PIDN:CAA34128.1; PID:g37663

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cell

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43921

A:Molecule type: mRNA

A:Residues: 208-440;1094-1385;1910-2246 <YAO>

R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.

Brain Res. Bull. 22, 67-70, 1989

A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilag

A:Reference number: A60979; MUID:89229983; PMID:2469524

A:Accession: A60979

A:Molecule type: protein

A:Residues: 171-210;289-303 <BIG>

R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.

J. Biol. Chem. 264, 5981-5987, 1989

A:Title: Isolation and partial characterization of a glial hyaluronate-binding prote

A:Reference number: A30358; MUID:89174663; PMID:2466833

A:Accession: A30358

A:Molecule type: protein

A:Residues: 24-50;80-87,'D',89-119;128-155;167-218;229-259,'IR';261-268;277-283,'G',

R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.

J. Biol. Chem. 262, 13120-13125, 1987

A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-

A:Reference number: A29348; MUID:88007514; PMID:2820964

A:Accession: A29348

A:Molecule type: mRNA

A:Residues: 1725,'V',1727-2409 <KRU>

A:Cross-references: GB:J02814

R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.

J. Biol. Chem. 267, 23883-23887, 1992

A:Title: Isolation of a large aggregating proteoglycan from human brain.

A:Reference number: A45131; MUID:93054750; PMID:1429726

A:Contents: brain

A:Accession: A45131

A:Molecule type: protein

A:Residues: 21-22,'X',24-37 <PE2>

A:Experimental source: brain

A>Note: sequence extracted from NCBI backbone (NCBIP:118884)

R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.

Genomics 14, 845-851, 1992

A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human

A:Reference number: I54179; MUID:93122792; PMID:1478664

A:Accession: I54179

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 251-347 <RES>

A;Cross-references: GB:S52488; NID:g263313; PIDN:AAB24878.1; PID:g263314
 C;Genetics:
 A;Gene: GDB:CSPG2
 A;Cross-references: GDB:127873; OMIM:118661
 A;Map position: 5q12-5q14
 C;Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EGF homology
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-2409/Product: proteoglycan 24K core protein #status predicted <MAT>
 F:167-244/Domain: link protein repeat homology <LNK1>
 F:265-346/Domain: link protein repeat homology <LNK2>
 F:559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
 F:2106-2137/Domain: EGF homology <EG1>
 F:2144-2175/Domain: EGF homology <EG2>
 F:2182-2302/Domain: C-type lectin homology <LCH>
 F:2309-2365/Domain: complement factor H repeat homology <FHD>
 Query Match 87.5%; Score 49; DB 1; Length 2409;
 Best Local Similarity 80.0%; Pred. No. 3.7;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 || |||||:|
 Db 2128 CVPYSGDQC 2137
 || |||||:|
 RESULT 9
 T42389
 N;Alternate names: versican, splice form V0 - bovine
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
 A;Accession: T42389
 R;Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
 J. Biol. Chem. 273, 15758-15764, 1998
 A;Title: Versican V2 is a major extracellular matrix component of the mature bovine brain
 A;Reference number: 217954; MUID:98288320; PMID:9624174
 A;Accession: T42389
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-3381 <SCH>
 A;Cross-references: EMBL:AF060456; NID:g3253299; PID:g3253300; PIDN:AAC24358.1
 C;Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin
 C;Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-3381/Product: versican, splice form V0 #status predicted <MAT>
 F:57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,24
 Query Match 87.5%; Score 49; DB 2; Length 3381;
 Best Local Similarity 80.0%; Pred. No. 5.1;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CVIGYSGDXC 10
 || |||||:|
 Db 3100 CVPYSGDQC 3109
 || |||||:|
 RESULT 10
 PC4415
 ErbB kinase activator beta, brain and thymus - rat (fragment)
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 10-Dec-1997 #sequence_revision 10-Dec-1997 #text_change 02-Aug-2002
 A;Accession: PC4415
 R;Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Miyag
 J. Biochem. 122, 675-680, 1997
 A;Title: A novel brain-derived member of the epidermal growth factor family that interac
 A;Reference number: JC5700; MUID:98006324; PMID:9348101
 A;Accession: PC4415
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-57 <HTG>
 A;Cross-references: DDBJ:D89997; NID:g2605633; PIDN:BAA23346.1; PID:g2605634
 A;Experimental source: PC-12 cell
 C;Comment: This protein is a member of the epidermal growth factor family. It is functio

ating the differentiation of MDA-MB-453 cells.

C;Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology
 F:1-25/Domain: EGF homology (fragment) <EGF>

Query Match 82.1%; Score 46; DB 2; Length 57;
 Best Local Similarity 60.0%; Pred. No. 0.34;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
 | :|||:|
 Db 16 CVPYGTGDC 25
 | :|||:|

RESULT 11

EGRT

epidermal growth factor precursor - rat

N;Alternate names: urogastrone precursor

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 14-Aug-1998 #text_change 18-Jun-1999

C;Accession: I52995; S05074; S01974; A25425; S18419; S08288

R;Saggi, S.J.; Safirstein, R.; Price, P.M.

DNA Cell Biol. 11, 481-487, 1992

A;Title: Cloning and Sequencing of the Rat Preproepidermal Growth Factor cDNA: Compa

A;Reference number: I52995; MUID:92398779; PMID:1524680

A;Accession: I52995

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1133 <RES>

A;Cross-references: EMBL:U04842; NID:g440236; PIDN:AAB60436.1; PID:g440237

R;Simpson, R.

submitted to the EMBL Data Library, August 1988

A;Reference number: S05074

A;Accession: S05074

A;Molecule type: mRNA

A;Residues: 'W',966,'RWL',970-1023,'NW',1026-1108,'SGAGVSSGPQWFFVLE',1126,'HQ' <S>

A;Cross-references: EMBL:X12748

R;Dorow, D.S.; Simpson, R.J.

Nucleic Acids Res. 16, 9338, 1988

A;Title: Cloning and sequence analysis of a cDNA for rat epidermal growth factor.

A;Reference number: S01974; MUID:89016634; PMID:3262867

A;Accession: S01974

A;Molecule type: mRNA

A;Residues: 'W',966,'RWL',970-1023,'NW',1026-1108 <DOR>

A;Cross-references: EMBL:X12748

R;Simpson, R.J.; Smith, J.A.; Moritz, R.L.; O'Hare, M.J.; Rudland, P.S.; Morrison, J.

Eur. J. Biochem. 153, 629-637, 1985

A;Title: Rat epidermal growth factor: complete amino acid sequence.

A;Reference number: A25425; MUID:86081810; PMID:3000782

A;Accession: A25425

A;Molecule type: protein

A;Residues: 974-1021 <SIG>

R;Nishi, N.; Shimizu, C.; Okutani, T.; Kagawa, Y.; Takasuga, H.; Suno, M.; Wada, F.

Biochim. Biophys. Acta 1095, 268-275, 1991

A;Title: Rat prostatic growth factors: Purification and characterization of high an

A;Reference number: S18419; MUID:92069070; PMID:1958699

A;Accession: S18419

A;Status: preliminary

A;Molecule type: protein

A;Residues: 974-1021 <NIS>

R;Nexo, E.; Jorgensen, P.E.; Thim, L.; Roepstorff, P.

Biochim. Biophys. Acta 1037, 388-393, 1990

A;Title: Purification and characterization of a low and a high molecular weight form

A;Reference number: S08288; MUID:90181442; PMID:2310752

A;Accession: S08288

A;Molecule type: protein

A;Residues: 974-1024 <NEX>

C;Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentia

gastrointestinal cell proliferation.

C;Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in s

C;Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound p

C;Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-con

C;Keywords: duplication; growth factor; mitogen; tandem repeat; transmembrane prote-n

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-1133/Product: epidermal growth factor proprotein, membrane-bound form #status predicted
F:22-1035/Region: EGF precursor long repeat <EXT>
F:44-480/Domain: LDL receptor WYTD-containing repeat homology <YW01>
F:47-86/Domain: LDL receptor WYTD-containing repeat homology <YW02>
F:87-128/Domain: LDL receptor WYTD-containing repeat homology <YW03>
F:129-170/Domain: LDL receptor WYTD-containing repeat homology <YW04>
F:171-212/Domain: LDL receptor WYTD-containing repeat homology <YW05>
F:213-257/Domain: LDL receptor WYTD-containing repeat homology <YW06>
F:258-302/Domain: LDL receptor WYTD-containing repeat homology <YW07>
F:319-355/Domain: EGF homology #status atypical <EG1>
F:361-396/Domain: EGF homology <EG2>
F:402-437/Domain: EGF homology <EG3>
F:440-477/Domain: EGF homology <EG4>
F:482-958/Region: EGF precursor long repeat <LR2>
F:485-525/Domain: LDL receptor WYTD-containing repeat homology <YW07>
F:526-568/Domain: LDL receptor WYTD-containing repeat homology <YW08>
F:569-611/Domain: LDL receptor WYTD-containing repeat homology <YW09>
F:612-655/Domain: LDL receptor WYTD-containing repeat homology <YW10>
F:656-696/Domain: LDL receptor WYTD-containing repeat homology <YW11>
F:697-739/Domain: LDL receptor WYTD-containing repeat homology <YW12>
F:747-782/Domain: EGF homology <EG5>
F:839-872/Domain: EGF homology <EG6>
F:878-914/Domain: EGF homology <EG7>
F:920-955/Domain: EGF homology <EG8>
F:974-1024/Product: epidermal growth factor #status experimental <MAT>
F:979-1015/Domain: EGF homology <EG9>
F:1036-1060/Domain: transmembrane #status predicted <TMM>
F:1061-1133/Domain: intracellular #status predicted <INT>
F:342-355,361-372,368-381,383-396,402-413,409-422,424-437,440-452,448-462,464-477,747-758-1015/Disulfide bonds: #status predicted

Query Match 82.1%; Score 46; DB 1; Length 1133;

Best Local Similarity 70.0%; Pred. No. 6;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :||

Db 1006 CVIGYIGERC 1015

RESULT 12

T42218

slit-1 protein homolog - rat

N:Alternate names: MEGF4 protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002

C:Accession: T42218

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs

A:Reference number: Z14126; MUID:98360089; PMID:9693030

A:Accession: T42218

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1531 <NAK>

A:Cross-references: EMBL:AB011530; NID:g3449289; PIDN:BAA32460.1; PID:g3449290

A:Experimental source: strain Sprague-Dawley; brain

C:Genetics:

A:Gene: MEGF4

C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein

Query Match

Best Local Similarity 82.1%; Score 46; DB 2; Length 1531;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :||

Db 1071 CVPGYTGDC 1080

RESULT 13

A31246

neurogenic protein Delta precursor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 02-Aug-2002
C:Accession: A31246
R:Kopczynski, C.C.; Alton, A.K.; Rechtel, K.; Kooh, P.J.; Muskavitch, M.A.T.
Genes Dev. 2, 1733-1735, 1988
A:Title: Delta, a Drosophila neurogenic gene, is transcriptionally complex and encoded
A:Reference number: A31246; MUID:89196890; PMID:3149249
A:Accession: A31246
A:Molecule type: mRNA
A:Residues: 1-832 <KOP>
A:Cross-references: GB:Y00222
C:Genetics:

A:Gene: FlyBase:DL

A:Cross-references: FlyBase:FBgn0000463

C:Superfamily: neurogenic protein delta; EGF homology

F:295-328/Domain: EGF homology <EGX1>

F:422-450/Domain: EGF homology <EGF1>

F:457-488/Domain: EGF homology <EGF>

F:533-564/Domain: EGF homology <EGF3>

Query Match 80.4%; Score 45; DB 2; Length 832;

Best Local Similarity 70.0%; Pred. No. 6.6;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :||

Db 319 CAPGYSGDDC 328

RESULT 14

S19087

gene Delta protein precursor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002

C:Accession: S19087

R:Muskavitch, M.A.T.

submitted to the EMBL Data Library, June 1991

A:Reference number: S19087

A:Accession: S19087

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-833 <MUS>

A:Cross-references: EMBL:Y00222

C:Genetics:

A:Gene: FlyBase:DL

A:Cross-references: FlyBase:FBgn0000463

C:Superfamily: neurogenic protein delta; EGF homology

F:335-371/Domain: EGF homology <EGF1>

F:378-415/Domain: EGF homology <EGX1>

F:457-488/Domain: EGF homology <EGF>

F:533-564/Domain: EGF homology <EGF3>

Query Match 80.4%; Score 45; DB 2; Length 833;

Best Local Similarity 70.0%; Pred. No. 6.6;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10

||||| :||

Db 319 CAPGYSGDDC 328

RESULT 15

S00670

neurogenic repetitive locus delta protein precursor - fruit fly (Drosophila melanogaster)

N:Alternate names: gene DL protein

C:Species: Drosophila melanogaster

C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 02-Aug-2002

C:Accession: S00670; A26637

R:Vaessin, H.; Bremer, K.A.; Knust, E.; Campos-Ortega, J.A.

EMBO J. 6, 3431-3440, 1987

A:Title: The neurogenic gene Delta of Drosophila melanogaster is expressed in neurog

A:Reference number: S00670

A:Accession: S00670

A:Molecule type: mRNA
A:Residues: 1-880 <VAE>
A:Cross-references: EMBL:X06289; NID:g7852; PID:g7853
R:Knust, E.; Dietrich, U.; Tepass, U.; Bremer, K.A.; Weigelt, D.; Vaessin, H.; Campos-Ortega
EMBO J. 6, 761-766, 1987
A:Title: EGF homologous sequences encoded in the genome of Drosophila melanogaster, and
A:Reference number: A91081; MUID:87218537; PMID:3107986
A:Accession: A26637
A:Molecule type: mRNA
A:Residues: 422-436, 'ET', 439-458, 'A', 460-489, 'T', 491-621 <KNU>
A:Cross-references: GB:X05140; NID:g7851; PIDN:CAA28786.1; PID:g929563
C:Genetics:
A:Gene: Delta; DL
A:Cross-references: FlyBase:FBgn0000463
C:Superfamily: neurogenic protein delta; EGF homology
C:Keywords: transmembrane protein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-88/Product: neurogenic repetitive locus delta protein #status predicted <MAT>
F:457-488/Domain: EGF homology <EGF1>
F:533-564/Domain: EGF homology <EGF2>

Query Match 80.4%; Score 45; DB 2; Length 880;
Best Local Similarity 70.0%; Pred. No. 7;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 CVIGYSGDXC 10
Db 319 CAPGYSGDDC 328

Search completed: July 2, 2003, 07:26:26
Job time : 12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:02 ; Search time 10.3333 Seconds
(without alignments)
40.138 Million cell updates/sec

Title: US-09-673-785D-9
Perfect score: 56
Sequence: 1 CVIGYSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1217	1	EGF_MOUSE
2	51	91.1	2139	1	CRB_DROME
3	49	87.5	862	1	PGCV_MACNE
4	49	87.5	3381	1	PGCV_BOVIN
5	49	87.5	3396	1	PGCV_HUMAN
6	48	85.7	2920	1	CLR2_MOUSE
7	46	82.1	756	1	NRG2_MOUSE
8	46	82.1	1133	1	EGF_RAT
9	45	80.4	833	1	DL_DROME
10	45	80.4	883	1	PGCB_RAT
11	45	80.4	1207	1	EGF_HUMAN
12	45	80.4	2319	1	NTC3_RAT
13	45	80.4	2738	1	PGCV_RAT
14	45	80.4	3358	1	PGCV_MOUSE
15	44	78.6	1064	1	FBP1_STRPU
16	44	78.6	2470	1	NTC2_MOUSE
17	44	78.6	2471	1	NTC2_HUMAN
18	44	78.6	2471	1	NTC2_RAT
19	44	78.6	2476	1	ZAN_PIG
20	44	78.6	4289	1	TENX_HUMAN
21	43	76.8	2321	1	NTC3_HUMAN
22	43	76.8	3672	1	LM22_CAEEL
23	42	75.0	53	1	EGF_PIG
24	42	75.0	57	1	Y402_ARCFU
25	42	75.0	473	1	FP2_MVTGA
26	42	75.0	883	1	PGCB_MOUSE
27	42	75.0	912	1	PGCB_BOVIN
28	42	75.0	1408	1	SERR_DROME
29	42	75.0	2318	1	NTC3_MOUSE
30	41	73.2	597	1	BP10_PARLI
31	41	73.2	1049	1	ADP1_YEAST
32	41	73.2	1192	1	LMG2_MOUSE
33	41	73.2	2437	1	NTC1_BRARE

RESULT 1
EGF_MOUSE
ID EGF_MOUSE STANDARD; PRT; 1217 AA.
AC P01132;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].
DE GN EGF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN [2]
RX MEDLINE=83223630; PubMed=6602382;
RA Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Fong N.M., Selby M., Rutter W.J., Bell G.I.;
RT "Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.";
RL Science 221:236-240(1983).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=83219309; PubMed=6304537;
RA Gray A., Dull T.J., Ullrich A.;
RT "Nucleotide sequence of epidermal growth factor cDNA predicts a Nature 303:722-725(1983).
RN [3]
RX SEQUENCE OF 977-1029.
RX MEDLINE=73048516; PubMed=4636327;
RA Savage C.R. Jr., Inagami T., Cohen S.;
RT "The primary structure of epidermal growth factor.";
RN [4]
RX J. Biol. Chem. 247:7612-7621(1972).
RN [5]
RX DISULFIDE BONDS.
RX MEDLINE=74025498; PubMed=4750422;
RA Savage C.R. Jr., Hash J.H., Cohen S.;
RT "Epidermal growth factor. Location of disulfide bonds.";
RN [5]
RX J. Biol. Chem. 248:7669-7672(1973).
RN [6]
RX STRUCTURE BY NMR OF 977-1029.
RX MEDLINE=92118798; PubMed=1731873;
RA Montellone G.T., Wuethrich K., Burgess A.W., Nice E.C., Wagner G., Gibson K.D., Scheraga H.A.;
RT "Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints.";
RN [6]
RX Biochemistry 31:236-249(1992).
RX STRUCTURE BY NMR OF 977-1029.
RX MEDLINE=93075811; PubMed=1445923;
RA Kohda D., Inagaki F.;
RT "Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions.";
RT

ALIGNMENTS

34 40 71.4 427 1 MFGM_RAT
35 40 71.4 463 1 MFGM_MOUSE
36 40 71.4 723 1 DLL1_HUMAN
37 40 71.4 1213 1 JAG3_BRARE
38 40 71.4 1218 1 JAG1_HUMAN
39 40 71.4 1218 1 JAG1_MOUSE
40 40 71.4 1219 1 JAG1_RAT
41 40 71.4 1242 1 JAG1_BRARE
42 40 71.4 1429 1 L112_CAEEL
43 40 71.4 2531 1 NTC1_MOUSE
44 40 71.4 2531 1 NTC1_RAT
45 40 71.4 2703 1 NOTC_DROME

P70490 rattus norv
P21956 mus musculu
Q00548 homo sapien
Q90y54 brachydanio
P78504 homo sapien
Q9qxx0 mus musculu
Q63722 rattus norv
Q90y57 brachydanio
P14585 caenorhabdi
Q01705 mus musculu
Q07008 rattus norv
P07207 drosophila

Biochemistry 31:11928-11939(1992).

[7]

RP STRUCTURE BY NMR OF 980-1024.

RA MEDLINE-99180407; PubMed=10082370;

RA Barnham K.J., Torres A.M., Alewood D., Alewood P.F., Domagala T.,

RA Nice E.C., Norton R.S.;

RT "Role of the 6-20 disulfide bridge in the structure and activity of

RT epidermal growth factor.";

RL Protein Sci. 7:1738-1749(1998).

CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS

CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME

CC FIBROBLASTS IN CELL CULTURE.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.

CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 1134

CC TO 1168 DUE TO A FRAMESHIFT.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; J00380; AAA37539.1; .

DR EMBL; V00741; CAA24115.1; ALT_FRAME.

DR EMBL; V00741; CAA24116.1; .

DR PIR; A01387; EGMMSG.

DR PDB; 1EGF; 31-JAN-94.

DR PDB; 3EGF; 31-JAN-94.

DR PDB; 1EPG; 31-JAN-94.

DR PDB; 1EPH; 31-JAN-94.

DR PDB; 1EPI; 31-JAN-94.

DR PDB; 1EPJ; 31-JAN-94.

DR PDB; 1A3P; 29-JUL-98.

DR MGD; MGI:95290; Egf.

DR InterPro; IPR000152; Asx_hydroxyl.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001336; EGF.1.

DR InterPro; IPR001881; EGF.Ca.

DR InterPro; IPR000033; Ldl_receptor_rep.

DR Pfam; PF00008; EGF; 8.

DR Pfam; PF00058; ldl_recept_b; 7.

DR PRINTS; PR00009; EGF.TGF.

DR SMART; SM00179; EGF.CA; 2.

DR SMART; SM00001; EGF like; 7.

DR SMART; SM00135; LY; 9.

DR PROSITE; PS00010; ASX_HYDROXYL; 3.

DR PROSITE; PS00022; EGF_1; 1.

DR PROSITE; PS01186; EGF_2; 6.

DR PROSITE; PS01187; EGF.CA; 3.

KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;

KW Signal; 3D-structure.

FT SIGNAL 1 28

FT CHAIN 29 1217

FT CHAIN 977 1029

FT DOMAIN 29 1038

FT TRANSMEM 1039 1058

FT DOMAIN 1059 1217

FT DOMAIN 327 361

FT DOMAIN 362 402

FT DOMAIN 403 443

FT DOMAIN 441 483

FT DOMAIN 747 787

FT DOMAIN 838 876

FT DOMAIN 877 918

FT DOMAIN 919 959

FT DOMAIN 978 1019

FT DOMAIN 366 377

FT DISULFID 373 386

FT DISULFID 388 401

FT DISULFID 407 418

POTENTIAL.

PRO-EPIDERMAL GROWTH FACTOR.

EPIDERMAL GROWTH FACTOR.

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOPLASMIC (POTENTIAL).

EGF-LIKE 1 (INCOMPLETE).

EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

EGF-LIKE 3.

EGF-LIKE 4.

EGF-LIKE 5.

EGF-LIKE 6.

EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).

EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).

EGF-LIKE 9.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

FT DISULFID 414 427

FT DISULFID 429 442

FT DISULFID 445 457

FT DISULFID 453 467

FT DISULFID 469 482

FT DISULFID 751 762

FT DISULFID 773 786

FT DISULFID 778 791

FT DISULFID 842 853

FT DISULFID 847 862

FT DISULFID 864 875

FT DISULFID 881 895

FT DISULFID 888 904

FT DISULFID 906 917

FT DISULFID 923 936

FT DISULFID 930 945

FT DISULFID 947 958

FT DISULFID 982 996

FT DISULFID 990 1007

FT DISULFID 1009 1018

FT DOMAIN 1024 1029

FT CARBOHYD 111 111

FT CARBOHYD 410 410

FT CARBOHYD 810 810

FT CARBOHYD 944 944

FT CONFLICT 790 790

FT CONFLICT 1048 1048

FT STRAND 995 997

FT STRAND 1006 1008

FT STRAND 1010 1010

FT TURN 1011 1012

FT STRAND 1013 1014

FT STRAND 1020 1021

SQ SEQUENCE 1217 AA; 133144 MW; A9C7F3D512F82873 CRC64;

NOT REQUIRED FOR FULL BIOLOGICAL ACTIVITY

IN VIVO.

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

D -> Y (IN REF. 2).

A -> S (IN REF. 2).

Query Match 100.0%; Score 56; DB 1; Length 1217;

Best Local Similarity 90.0%; Pred. No. 0.027;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CVIGSGDNC 10

Db 1009 CVIGSGDNC 1018

|||||||

RESULT 2

CRB_DROME STANDARD; PRT; 2139 AA.

AC P10040;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Crumbs protein precursor (95F).

GN CRB.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R; TISSUE=Embryo;

RX MEDLINE=90263104; PubMed=2344615;

RA Tepass U., Theres C., Knust E.;

RT "Crumbs encodes an EGF-like protein expressed on apical membranes of

RT Drosophila epithelial cells and required for organization of

RT epithelia.";

RL Cell 61:787-799(1990).

RN [2]

RP SEQUENCE OF 1663-1955 FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=87218537; PubMed=3107986;

RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,

FT DISULFID 1768 1783 BY SIMILARITY.
 FT DISULFID 1785 1794 BY SIMILARITY.
 FT DISULFID 1801 1812 BY SIMILARITY.
 FT DISULFID 1806 1821 BY SIMILARITY.
 FT DISULFID 1823 1832 BY SIMILARITY.
 FT DISULFID 1839 1850 BY SIMILARITY.
 FT DISULFID 1844 1859 BY SIMILARITY.
 FT DISULFID 1861 1870 BY SIMILARITY.
 FT DISULFID 1878 1889 BY SIMILARITY.
 FT DISULFID 1883 1903 BY SIMILARITY.
 FT DISULFID 1905 1914 BY SIMILARITY.
 FT DISULFID 1919 1930 BY SIMILARITY.
 FT DISULFID 1924 1939 BY SIMILARITY.
 FT DISULFID 1941 1950 BY SIMILARITY.
 FT DISULFID 1957 1968 BY SIMILARITY.
 FT DISULFID 1962 1977 BY SIMILARITY.
 FT DISULFID 1979 1988 BY SIMILARITY.
 FT DISULFID 1995 2008 BY SIMILARITY.
 FT DISULFID 2002 2017 BY SIMILARITY.
 FT DISULFID 2019 2028 BY SIMILARITY.
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 91.1%; Score 51; DB 1; Length 2139;
 Best Local Similarity 70.0%; Pred. No. 0.37;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
 Db 571 CAVIGSGDRC 580

RESULT 3
 PGCV_MACNE STANDARD; PRT; 862 AA.
 AC Q28858: Q28859; Q28860;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE versican core protein (Large fibroblast proteoglycan) (Chondroitin sulfate proteoglycan core protein 2) (Fragments).
 GN CSPG2.
 OS Macaca nemestrina (Pig-tailed macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9545;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Aortic smooth muscle;
 RX MEDLINE=95005762; PubMed=7921538;
 RA Yao L.Y., Moody C., Schoenher E., Wight T.N., Sandell L.J.;
 RT "Identification of the proteoglycan versican in aorta and smooth muscle cells by DNA sequence analysis, in situ hybridization and immunohistochemistry."
 RL Matrix Biol. 14:213-225(1994).
 CC -1- FUNCTION: May play a role in intercellular signaling and in connecting cells with the extracellular matrix. May take part in the regulation of cell motility, growth and differentiation. Binds hyaluronan.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development (By similarity).
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.

CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S72412; AAA65593.2; -;
 DR EMBL; S72413; AAA65594.2; -;
 DR EMBL; S72414; AAA65595.2; -;
 DR HSSP; P01132; IEGF.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF-2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR PRINTS; PR00010; EGFBL00D.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00001; EGF_Like; 1.
 DR SMART; SM00445; LINK; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS01241; LINK; 1.
 DR PROSITE; PS00615; C-TYPE-LECTIN_1; PARTIAL.
 DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Repeat;
 KW EGF-like domain; Calcium.
 FT NON_TER 1 1
 FT DOMAIN <1 37 LINK 1.
 FT DOMAIN 58 139 LINK 2.
 FT DOMAIN 141 >233 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
 FT NON_CONS 233 234 GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN,
 FT DOMAIN <234 >525 SIMILAR TO CHONDROITIN SULFATE ATTACHMENT SITE IN COLLAGEN TYPE IX (BY SIMILARITY).
 FT NON_CONS 525 526 EGF-LIKE 1.
 FT DOMAIN 718 754 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 756 792 C-TYPE LECTIN.
 FT DISULFID 63 138 BY SIMILARITY.
 FT DISULFID 87 108 BY SIMILARITY.
 FT DISULFID 722 733 BY SIMILARITY.
 FT DISULFID 727 742 BY SIMILARITY.
 FT DISULFID 744 753 BY SIMILARITY.
 FT DISULFID 760 771 BY SIMILARITY.
 FT DISULFID 765 780 BY SIMILARITY.
 FT DISULFID 782 791 BY SIMILARITY.
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 563 563 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 862 862
 SQ SEQUENCE 862 AA; 95583 MW; A5D5F6153A74BB39 CRC64;

Query Match 87.5%; Score 49; DB 1; Length 862;
 Best Local Similarity 80.0%; Pred. No. 0.35;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10

FT CARBOHYD 2074 2074 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2103 2103 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2263 2263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2290 2290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2356 2356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2623 2623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2641 2641 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2919 2919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3052 3052 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3354 3354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3364 3364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 349 349 P -> R (IN ISOFORM V1 AND V3).
 FT VARSPLIC 350 1336 MISSING (IN ISOFORM V1).
 FT VARSPLIC 1337 3074 MISSING (IN ISOFORM V2).
 FT VARSPLIC 350 3074 MISSING (IN ISOFORM V3).
 FT CONFLICT 25 25 MISSING (IN REF. 2).
 FT CONFLICT 51 51 N -> D (IN REF. 2).
 FT CONFLICT 89 89 Q -> D (IN REF. 2).
 FT CONFLICT 96 96 C -> R (IN REF. 2).
 FT CONFLICT 346 346 C -> R (IN REF. 2).
 SQ SEQUENCE 3381 AA; 369984 MW; F09716FA7778D459 CRC64;
 Query Match 87.5%; Score 49; DB 1; Length 3381;
 Best Local Similarity 80.0%; Pred. No. 1.3;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CVTGYSGDXC 10
 DB 3100 CVPGYSGDRC 3109
 RESULT 5
 PGCV HUMAN
 ID PGCV HUMAN STANDARD; PRT: 3396 AA.
 AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
 DE hyaluronate-binding protein) (GHAP).
 GN GSPG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM V0).
 RX MEDLINE-95105188; PubMed-7528742;
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
 RT "Characterization of the complete genomic structure of the human
 RT versican gene and functional analysis of its promoter.";
 RL J. Biol. Chem. 269:32999-33008(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V1).
 RC TISSUE=Placenta;
 RX MEDLINE-90059882; PubMed-2583089;
 RA Zimmermann D.R., Ruoslahti E.;
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";
 RL EMBO J. 8:2975-2981(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM V2).
 RC TISSUE=Glial tumor;
 RX MEDLINE-95105187; PubMed-7806529;
 RA Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "A novel glycosaminoglycan attachment domain identified in two
 RT alternative splice variants of human versican.";
 RL J. Biol. Chem. 269:32992-32998(1994).
 RN [4]
 RP SEQUENCE OF 2711-3396 FROM N.A.
 RC TISSUE=Lung fibroblast;
 RX MEDLINE-88007514; PubMed-2820964;
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;

RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 RT lectin-like and growth factor-like sequences.";
 RL J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE-93122792; PubMed-1478664;
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McPherson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 RT human chromosome 5 (9q12-5q14).";
 RL Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RC TISSUE=Brain;
 RX MEDLINE-95181355; PubMed-7876137;
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RT without a chondroitin sulfate attachment in region in mouse and human
 RT tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [7]
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
 RC TISSUE=Arctic smooth muscle;
 RX MEDLINE-99327053; PubMed-10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wight T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [8]
 RP PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE-89174663; PubMed-2466833;
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
 RT "Isolation and partial characterization of a glial
 RT hyaluronate-binding protein.";
 RL J. Biol. Chem. 264:5981-5987(1989).
 RN [9]
 RP TISSUE SPECIFICITY OF ISOFORMS.
 RX MEDLINE-96213482; PubMed-8627343;
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -1- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
 CC V2, V3 and Vint; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; V2 is restricted to normal brain
 CC and gliomas; V3 is found in all these tissues except
 CC medulloblastomas.
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U16306; AAA65018.1; -;
 DR EMBL; X15998; CAA34128.1; -;
 DR EMBL; S52488; AAB24878.1; -;

DR EMBL; U26555; AAA67565.1; .
DR EMBL; D32039; BAA06801.1; .
DR EMBL; J02814; AAA36437.1; .
DR EMBL; AF084545; AAD48545.1; .
DR PIR; S06014; S06014.
DR PIR; A29348; A29348.
DR PIR; A30358; A30358.
DR HSSP; P01132; 1EGF.
DR Genew; HGNC:2464; CSPG2.
DR MIM; 118661; .
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_C; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; xlink; 2.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRODOM; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF-like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
DR Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 3396
FT DOMAIN 37 137
FT DOMAIN 167 244
FT DOMAIN 265 346
FT DOMAIN 348 1335
FT DOMAIN 1336 3089
FT DOMAIN 3089 3125
FT DOMAIN 3127 3163
FT DOMAIN 3176 3290
FT DOMAIN 3295 3353
FT DISULFID 44 130
FT DISULFID 172 243
FT DISULFID 196 217
FT DISULFID 270 345
FT DISULFID 294 315
FT DISULFID 3093 3104
FT DISULFID 3098 3113
FT DISULFID 3115 3124
FT DISULFID 3131 3142
FT DISULFID 3136 3151
FT DISULFID 3153 3162
FT DISULFID 3169 3180
FT DISULFID 3197 3289
FT DISULFID 3265 3281
FT DISULFID 3296 3339
FT DISULFID 3325 3352
FT CARBOHYD 57 57
FT CARBOHYD 330 330

FT CARBOHYD 615 615
FT CARBOHYD 782 782
FT CARBOHYD 809 809
FT CARBOHYD 1332 1332
FT CARBOHYD 1398 1398
FT CARBOHYD 1442 1442
FT CARBOHYD 1468 1468
FT CARBOHYD 1663 1663
FT CARBOHYD 1898 1898
FT CARBOHYD 2179 2179
FT CARBOHYD 2272 2272
FT CARBOHYD 2280 2280
FT CARBOHYD 2360 2360
FT CARBOHYD 2385 2385
Query Match 87.5%; Score 49; DB 1; Length 3396;
Best Local Similarity 80.0%; Pred.No. 1.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CVIGVSGDXC 10
Db 3115 CVPVSGDQC 3124
RESULT 6
CLR2_MOUSE STANDARD; PRT: 2920 AA.
ID CLR2_MOUSE Q922R4; Q99K26;
AC Q9ROM0; Q922R4; Q99K26;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)
DE (mEmil).
GN CELSR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A. PubMed-10490098;
RA MEDLINE-94118630; PubMed-10490098;
RX Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
polarity under the control of frizzled.";
RL Cell 98:585-595(1999).
RN [2]
RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.
RX PubMed-10790539;
RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;
RT "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is
a candidate for the tipy (tip) lethal mutant on chromosome 9.";
RL Mamm. Genome 11:392-394(2000).
RN [3]
RP SEQUENCE OF 2014-2920 FROM N.A.
RC TISSUE-Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP DEVELOPMENTAL STAGE.
RX PubMed-11850187;
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celsr (Flamingo) genes in the
mouse.";
RL Mech. Dev. 112:157-160(2002).
CC -|- FUNCTION: Receptor that may have an important role in cell/cell
signaling during nervous system formation.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- DEVELOPMENTAL STAGE: Predominantly expressed in the developing
CNS, the emerging dorsal root ganglia and cranial ganglia. In the
CNS, expression is uniform along the rostrocaudal axis. During
gastrulation, it is expressed within the anterior neural ectoderm.
CC At E10, expression is strong in the ventricular zones (VZ) in all

sectors of the brain, and lower in the marginal zones (MZ).
 Between E12 and E15, expression is prominent in the brain. It is
 strong in VZ, lower in MZ, except in telecephalic MZ where it is
 predominant. The intensity is higher in all VZ, and lower in
 differentiating fields than in VZ, except in the cerebral
 hemispheres, and to a lesser extent in the tectum and cerebellum.
 A weak expression is also observed in the fetal lungs, kidney and
 epithelia. In the newborn and postnatal stages, expression remains
 restricted to the VZ as well as in migrating and postmigratory
 cells throughout the brain.

- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.
 - SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
 - SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.
 - SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
 - SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 - SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.
 - SIMILARITY: CONTAINS 1 GPS DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AB028499; BAA84070.1; -;
 EMBL; AF031573; AAC68837.1; -;
 EMBL; BC005499; AAH05499.1; -;
 HSP; P00740; 1EDM.
 MGD; MGI:1858235; Celar2.
 InterPro: IPR000152; Asx_hydroxyl.
 InterPro: IPR002126; Cadherin.
 InterPro: IPR000561; EGF-like.
 InterPro: IPR000742; EGF-2.
 InterPro: IPR000832; GPCR_secretin.
 InterPro: IPR001879; hormn_receptor.
 InterPro: IPR002049; Laminin_EGF.
 InterPro: IPR001791; Laminin_G.
 InterPro: IPR000203; PKD_cys_rich.
 Pfam; PF00002; 7tm_2; 1.
 Pfam; PF00028; cadherin; 9.
 Pfam; PF00008; EGF; 5.
 Pfam; PF01825; GPS; 1.
 Pfam; PF02793; HRM; 1.
 Pfam; PF00054; laminin_G; 2.
 PRINTS; PR00205; CADHERIN.
 PRINTS; PR00011; EGF_LAMININ.
 PRINTS; PR00249; GPCRSECRETIN.
 SMART; SM00112; CA; 9.
 SMART; SM00180; EGF_Lam; 1.
 SMART; SM00001; EGF_like; 6.
 SMART; SM00303; GPS; 1.
 SMART; SM00008; HormR; 1.
 SMART; SM00282; LamG; 2.
 PROSITE; PS00010; ASX_HYDROXYL; 1.
 PROSITE; PS00232; CADHERIN_1; 6.
 PROSITE; PS00268; CADHERIN_2; 9.
 PROSITE; PS00022; EGF_1; 6.
 PROSITE; PS01186; EGF_2; 2.
 PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
 PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
 PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
 PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
 PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 PROSITE; PS02021; GPS; 1.
 PROSITE; PS00025; LAM_G_DOMAIN; 2.
 G-protein coupled receptor; Transmembrane; Glycoprotein;
 EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
 Developmental protein; Hydroxylation; Signal.
 SIGNAL 1 31 POTENTIAL.
 CHAIN 32 2920 CADHERIN EGF LAG SEVEN-PASS G-TYPE
 RECEPTOR 2.

FT	DOMAIN	32	2381	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2382	2402	1 (POTENTIAL).
FT	DOMAIN	2403	2414	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2415	2434	2 (POTENTIAL).
FT	DOMAIN	2435	2439	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2440	2460	3 (POTENTIAL).
FT	DOMAIN	2461	2481	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2482	2502	4 (POTENTIAL).
FT	DOMAIN	2503	2519	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2520	2540	5 (POTENTIAL).
FT	DOMAIN	2541	2564	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2565	2585	6 (POTENTIAL).
FT	DOMAIN	2586	2592	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2593	2613	7 (POTENTIAL).
FT	DOMAIN	2614	2920	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	182	289	CADHERIN 1.
FT	DOMAIN	290	399	CADHERIN 2.
FT	DOMAIN	400	506	CADHERIN 3.
FT	DOMAIN	507	611	CADHERIN 4.
FT	DOMAIN	612	713	CADHERIN 5.
FT	DOMAIN	714	816	CADHERIN 6.
FT	DOMAIN	817	922	CADHERIN 7.
FT	DOMAIN	923	1024	CADHERIN 8.
FT	DOMAIN	1029	1147	CADHERIN 9.
FT	DOMAIN	1229	1287	EGF-LIKE 1 (ATYPICAL).
FT	DOMAIN	1289	1319	EGF-LIKE 2, CALCIUM-BINDING.
FT	DOMAIN	1329	1367	EGF-LIKE 3, CALCIUM-BINDING.
FT	DOMAIN	1368	1572	LAMININ G-LIKE 1.
FT	DOMAIN	1575	1611	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	1615	1792	LAMININ G-LIKE 2.
FT	DOMAIN	1788	1830	EGF-LIKE 5, CALCIUM-BINDING.
FT	DOMAIN	1831	1868	EGF-LIKE 6, CALCIUM-BINDING.
FT	DOMAIN	1884	1923	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	1924	1956	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	2317	2369	GPS.
FT	DOMAIN	2744	2749	POLY-GLU.
FT	DISULFID	1293	1304	BY SIMILARITY.
FT	DISULFID	1298	1313	BY SIMILARITY.
FT	DISULFID	1315	1318	BY SIMILARITY.
FT	DISULFID	1333	1344	BY SIMILARITY.
FT	DISULFID	1338	1354	BY SIMILARITY.
FT	DISULFID	1356	1366	BY SIMILARITY.
FT	DISULFID	1379	1590	BY SIMILARITY.
FT	DISULFID	1584	1599	BY SIMILARITY.
FT	DISULFID	1601	1610	BY SIMILARITY.
FT	DISULFID	1792	1803	BY SIMILARITY.
FT	DISULFID	1798	1818	BY SIMILARITY.
FT	DISULFID	1820	1829	BY SIMILARITY.
FT	DISULFID	1833	1844	BY SIMILARITY.
FT	DISULFID	1838	1856	BY SIMILARITY.
FT	DISULFID	1858	1867	BY SIMILARITY.
FT	DISULFID	1888	1900	BY SIMILARITY.
FT	DISULFID	1890	1907	BY SIMILARITY.
FT	DISULFID	1909	1922	BY SIMILARITY.
FT	DISULFID	1925	1937	BY SIMILARITY.
FT	DISULFID	1927	1944	BY SIMILARITY.
FT	DISULFID	1946	1955	BY SIMILARITY.
FT	MOD_RES	1592	1592	HYDROXYLATION (POTENTIAL).
FT	CARBOHYD	486	486	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	558	558	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	702	702	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1037	1037	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1077	1077	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1183	1183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1502	1502	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1566	1566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1742	1742	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1828	1828	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1901	1901	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2025	2025	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2062	2062	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2324 2324 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2346 2346 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 2199 2199 V -> L (IN REF. 3).
 FT CONFLICT 2283 2283 A -> V (IN REF. 3).
 FT CONFLICT 2535 2535 S -> R (IN REF. 1).
 FT CONFLICT 2571 2571 L -> R (IN REF. 2).
 FT CONFLICT 2639 2639 Y -> S (IN REF. 3).

Query Match 85.7%; Score 48; DB 1; Length 2920;
 Best Local Similarity 70.0%; Pred. No. 1.7;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSDXC 10
 ||:|:| ||:|
 Db 1820 CVLGYGDNC 1829

RESULT 7
 NR2_MOUSE
 ID NR2_MOUSE STANDARD; PRT; 756 AA.
 AC P56974;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)
 (Divergent of neuregulin 1) (DON-1)].
 GN NRG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=97311398; PubMed=9168115;
 RA Caraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N.,
 Gassmann M., Lai C.;
 RT "Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine
 kinases.";
 RL Nature 387:512-516(1997).
 RN [2]
 RP SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).
 RC TISSUE=Choroid plexus;
 RX MEDLINE=97342638; PubMed=9199335;
 RA Woolfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,
 Wolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,
 Gearing D.P.;
 RT "Characterization of a neuregulin-related gene, Don-1, that is highly
 expressed in restricted regions of the cerebellum and hippocampus.";
 RL Mol. Cell. Biol. 17:4007-4014(1997).
 CC -|- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
 RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
 RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
 ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE
 HETERODIMERIZATION WITH THE EGF RECEPTOR.
 CC -|- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
 A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
 MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
 CC -|- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS: DON-1M, DON-1S/NRG2-5,
 NRG2-10 AND NRG2-16A (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE
 SPLICING.
 CC -|- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WITH LOWER
 LEVELS IN THE LUNG. IN THE CEREBELLUM, FOUND IN GRANULE AND
 PURKINJE CELLS.
 CC -|- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
 OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
 PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
 DIMERIZATION (BY SIMILARITY).
 CC -|- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
 DOMAIN (BY SIMILARITY).
 CC -|- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
 EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
 FORM (BY SIMILARITY).

CC -|- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
 SIMILARITY).
 CC -|- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -|- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
 DR HSP; Q12784; IHRE.
 DR MGD; MGI:1098246; Nrg2.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR002154; Neuregulin.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; Ig; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGC2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
 Transmembrane; Multigene family; Alternative splicing.
 FT PROPEP 1 19 BY SIMILARITY.
 FT CHAIN 20 756 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
 FT CHAIN 20 314 NEUREGULIN-2.
 FT DOMAIN 20 315 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 316 336 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
 FT DOMAIN 337 756 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 158 226 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 238 248 SER/THR-RICH.
 FT DOMAIN 249 290 EGF-LIKE.
 FT DOMAIN 627 633 POLY-PRO.
 FT DISULFID 165 219 BY SIMILARITY.
 FT DISULFID 253 267 BY SIMILARITY.
 FT DISULFID 261 278 BY SIMILARITY.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 280 280 C -> G (IN ISOFORM NRG2-10).
 FT VARSPLIC 281 756 MISSING (IN ISOFORM NRG2-10).
 FT VARSPLIC 282 330 VYTGDRCCQFAMVNFSLKGLFELKEAEELQKRVLTITGI
 CVALLVVG -> NGFFGRCLEKLPRLYMPDPKQSVLWDT
 PGTGVSSQSTSPSTLDLN (IN ISOFORM DON-1S).
 FT VARSPLIC 331 756 MISSING (IN ISOFORM DON-1S).
 FT VARSPLIC 282 307 VYTGDRCCQFAMVNFSLKGLFELKE -> NGFFGRCLEK
 LPLRLYMPDPKQK (IN ISOFORM DON-1M).
 SQ SEQUENCE 756 AA; 82213 MW; 51D85DC918BE678E CRC64;
 Query Match 82.1%; Score 46; DB 1; Length 756;
 Best Local Similarity 60.0%; Pred. No. 1.1;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSDXC 10
 | :|:|:|:|
 Db 280 CPVGYTGDRC 289

RESULT 8
 EGF_RAT
 ID EGF_RAT STANDARD; PRT; 1133 AA.
 AC P07522; Q63183;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal
 growth factor].
 DE EGF.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE-Kidney;
 RX MEDLINE=92398779; PubMed=1524680;
 RA Price P.M., Saggi S.J., Safirstein R.;
 RT "Cloning and sequencing of the rat preproepidermal growth factor
 RL cDNA: comparison with mouse and human sequences.";
 RNA DNA Cell Biol. 11:481-487(1992).
 [2]
 RN REVISIONS.
 RC TISSUE-Kidney;
 RA Price P.M.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE OF 974-1021.
 RX MEDLINE=86081810; PubMed=3000782;
 RA Simpson R.J., Smith J.A., Moritz R.L., O'Hare M.J., Rudland P.S.,
 RA Morrison J.R., Lloyd C.J., Grego B., Burgess A.W., Nice E.C.;
 RT "Rat epidermal growth factor: complete amino acid sequence. Homology
 RT with the corresponding murine and human proteins; isolation of a form
 RT truncated at both ends with full in vitro biological activity.";
 RL Eur. J. Biochem. 153:629-637(1985).
 [4]
 RN SEQUENCE OF 994-1108 FROM N.A.
 RX STRAIN-Sprague-Dawley; TISSUE-Kidney;
 RA MEDLINE=89016634; PubMed=3262867;
 RA Dorow D.S., Simpson R.J.;
 RT "Cloning and sequence analysis of a cDNA for rat epidermal growth
 RT factor";
 RL Nucleic Acids Res. 16:9338-9338(1988).
 CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
 CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
 CC FIBROBLASTS IN CELL CULTURE.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U04842; AAB60436.1; -;
 DR EMBL; X12748; CAA31241.1; -;
 DR PIR; A25425; EGRF.
 DR HSSP; P01132; LEGF.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001336; EGF_1.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR000033; Ldl_receptor_rep.
 DR Pfam; PF00008; EGF; 7.
 DR Pfam; PF00058; ldl_recept_b; 7.
 DR PRINTS; PR00009; EGF_TGF.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00001; EGF_like; 6.
 DR SMART; SM00135; LY; 8.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 6.
 DR PROSITE; PS01187; EGF_CA; 3.
 KW EGF-like domain; Repeat; growth factor; Transmembrane; Glycoprotein;
 KW Signal.
 FT SIGNAL
 FT CHAIN 22 1133 POTENTIAL.
 FT CHAIN 974 1026 PRO-EPIDERMAL GROWTH FACTOR.
 FT DOMAIN 22 1035 EPIDERMAL GROWTH FACTOR.
 FT TRANSMEM 1036 1057 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 1058 1133 POTENTIAL.
 FT DOMAIN 322 356 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 357 397 EGF-LIKE 1 (INCOMPLETE).
 FT DOMAIN 398 438 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 436 478 EGF-LIKE 3.
 FT DOMAIN 436 478 EGF-LIKE 4.

FT DOMAIN 743 783 EGF-LIKE 5.
 FT DOMAIN 835 873 EGF-LIKE 6.
 FT DOMAIN 874 915 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 916 956 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 975 1016 EGF-LIKE 9.
 FT DISULFID 361 372 BY SIMILARITY.
 FT DISULFID 368 381 BY SIMILARITY.
 FT DISULFID 383 396 BY SIMILARITY.
 FT DISULFID 402 413 BY SIMILARITY.
 FT DISULFID 409 422 BY SIMILARITY.
 FT DISULFID 424 437 BY SIMILARITY.
 FT DISULFID 440 452 BY SIMILARITY.
 FT DISULFID 448 462 BY SIMILARITY.
 FT DISULFID 464 477 BY SIMILARITY.
 FT DISULFID 747 758 BY SIMILARITY.
 FT DISULFID 754 767 BY SIMILARITY.
 FT DISULFID 769 782 BY SIMILARITY.
 FT DISULFID 839 850 BY SIMILARITY.
 FT DISULFID 844 859 BY SIMILARITY.
 FT DISULFID 861 872 BY SIMILARITY.
 FT DISULFID 878 892 BY SIMILARITY.
 FT DISULFID 885 901 BY SIMILARITY.
 FT DISULFID 903 914 BY SIMILARITY.
 FT DISULFID 920 933 BY SIMILARITY.
 FT DISULFID 927 942 BY SIMILARITY.
 FT DISULFID 944 955 BY SIMILARITY.
 FT DISULFID 979 993 BY SIMILARITY.
 FT DISULFID 987 1004 BY SIMILARITY.
 FT DISULFID 1006 1015 BY SIMILARITY.
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 955 955 C -> V.
 FT CONFLICT 1024 1025 KL -> NW (IN REF. 4).
 FT CONFLICT 1108 1108 Q -> S (IN REF. 4).
 SQ SEQUENCE 1133 AA; 124125 MW; C224A302E9578031 CRC64;
 Query Match 82.1%; Score 46; DB 1; Length 1133;
 Best Local Similarity 70.0%; Pred. No. 1.6;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 OY 1 CVIGISGDXC 10
 Db 1006 CVIGIGERC 1015
 ID DL_DROME STANDARD; PRT; 833 AA.
 AC P10041; Q9VDY2; Q99108;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus Delta protein precursor.
 GN DL OR CG3619.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RA Vaessin H., Bremer K.A., Knust E., Campos-Ortega J.A.;
 RT "The neurogenic gene Delta of Drosophila melanogaster is expressed in
 RT neurogenic territories and encodes a putative transmembrane protein
 RT with EGF-like repeats.";
 RL EMBO J. 6:3431-3440(1987).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-Oregon-R; TISSUE-Embryo;
RX MEDLINE=89196890; PubMed=3149249;
RA Kopczynski C.C., Altan A.K., Fechtel K., Kooh P.J., Muskavitch M.A.T.;
RT "Delta, a Drosophila neurogenic gene, is transcriptionally complex and
RT encodes a protein related to blood coagulation factors and epidermal
RT growth factor of vertebrates.";
RL Genes Dev. 2:1723-1735(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceinaker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananadis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [4]
RP SEQUENCE OF 422-621 FROM N.A.
RX TISSUE-Embryo;
RX MEDLINE=87218537; PubMed=3107986;
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D., Vaessin H.,
RA Campos-Ortega J.A.;
RT "EGF homologous sequences encoded in the genome of Drosophila
RT melanogaster, and their relation to neurogenic genes.";
RL EMBO J. 6:761-766(1987).
RN [5]
RP PATTERN OF TRANSCRIPTION, AND CHARACTERIZATION.
RX MEDLINE=91209246; PubMed=2128477;
RA Haenlin M., Kramatschek B., Campos-Ortega J.A.;
RT "The pattern of transcription of the neurogenic gene Delta of
RT Drosophila melanogaster.";
RL Development 110:905-914(1990).
CC -!- FUNCTION: ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM. DL
CC IS REQUIRED FOR THE CORRECT SEPARATION OF NEURAL AND EPIDERMAL
CC CELL LINESAGES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL AREAS WITH NEUROGENIC
CC ABILITIES, FOR EXAMPLE THE NEUROGENIC ECTODERM AND THE PRIMORDIA
CC OF THE SENSE ORGANS. LATER EXPRESSION IS RESTRICTED TO THOSE CELLS
CC THAT HAVE ADOPTED A NEURAL FATE.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC EXPRESSION IS HIGHEST EARLY IN EMBRYONIC DEVELOPMENT (STAGE 5) AND
CC REDUCES TO A LOW LEVEL DURING LARVAL STAGES.
CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
CC THE INNER PART OF EMBRIO IS ONE OF THE FIRST STEPS OF CNS
CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
CC NEUROGENIC GENES.
CC -!- MISCELLANEOUS: NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN
CC LEVEL, IT IS CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY
CC COMPLETE FOR BINDING WITH THE NOTCH PROTEIN.
CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DSL DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06289; CAA29617.1; -
CC EMBL: Y00222; CAA68369.1; -
CC EMBL: A0003725; AAF55657.1; -
CC PIR: A06637; S00670.
CC PIR: A26637; A26637.
CC HSP: P00740; IEDM.
CC FlyBase: FBgn0000463; DL.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR001774; DSL.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF-2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR001438; EGF_II.
CC InterPro: IPR002049; Laminin_EGF.
CC Pfam: PF000008; EGF; 9.
CC Pfam: PF01414; DSL; 1.
CC PRINTS: PR00010; EGFBL00D.
CC PRINTS: PR00011; EGF_LAMININ.
CC SMART: SM00051; DSL; 1.
CC SMART: SM00179; EGF_Ca; 4.
CC SMART: SM00001; EGF-like; 5.
CC PROSITE: PS00010; ASX_HYDROXYL; 3.
CC PROSITE: PS00022; EGF_1; 9.
CC PROSITE: PS01186; EGF-2; 9.
CC PROSITE: PS01187; EGF_Ca; 2.
KW Developmental protein; Differentiation; Neurogenesis; Repeat;
KW Transmembrane; EGF-like domain; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 833
FT DOMAIN 19 594
FT TRANSMEM 595 617
FT DOMAIN 618 833
FT DOMAIN 164 226
FT DOMAIN 227 258
FT DOMAIN 256 289
FT DOMAIN 291 329
FT DOMAIN 331 372
FT DOMAIN 374 416
FT DOMAIN 418 451
FT DOMAIN 453 489
FT DOMAIN 491 527
FT DOMAIN 529 565
FT DISULFID 231 240
FT DISULFID 235 246
FT DISULFID 248 257
FT DISULFID 260 271
FT DISULFID 266 277
FT DISULFID 279 288
FT DISULFID 295 307
FT DISULFID 301 317
FT DISULFID 319 328

CC -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC EXPRESSION IS HIGHEST EARLY IN EMBRYONIC DEVELOPMENT (STAGE 5) AND
CC REDUCES TO A LOW LEVEL DURING LARVAL STAGES.
CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
CC THE INNER PART OF EMBRIO IS ONE OF THE FIRST STEPS OF CNS
CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
CC NEUROGENIC GENES.
CC -!- MISCELLANEOUS: NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN
CC LEVEL, IT IS CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY
CC COMPLETE FOR BINDING WITH THE NOTCH PROTEIN.
CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DSL DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06289; CAA29617.1; -
CC EMBL: Y00222; CAA68369.1; -
CC EMBL: A0003725; AAF55657.1; -
CC PIR: A06637; S00670.
CC PIR: A26637; A26637.
CC HSP: P00740; IEDM.
CC FlyBase: FBgn0000463; DL.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR001774; DSL.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF-2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR001438; EGF_II.
CC InterPro: IPR002049; Laminin_EGF.
CC Pfam: PF000008; EGF; 9.
CC Pfam: PF01414; DSL; 1.
CC PRINTS: PR00010; EGFBL00D.
CC PRINTS: PR00011; EGF_LAMININ.
CC SMART: SM00051; DSL; 1.
CC SMART: SM00179; EGF_Ca; 4.
CC SMART: SM00001; EGF-like; 5.
CC PROSITE: PS00010; ASX_HYDROXYL; 3.
CC PROSITE: PS00022; EGF_1; 9.
CC PROSITE: PS01186; EGF-2; 9.
CC PROSITE: PS01187; EGF_Ca; 2.
KW Developmental protein; Differentiation; Neurogenesis; Repeat;
KW Transmembrane; EGF-like domain; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 833
FT DOMAIN 19 594
FT TRANSMEM 595 617
FT DOMAIN 618 833
FT DOMAIN 164 226
FT DOMAIN 227 258
FT DOMAIN 256 289
FT DOMAIN 291 329
FT DOMAIN 331 372
FT DOMAIN 374 416
FT DOMAIN 418 451
FT DOMAIN 453 489
FT DOMAIN 491 527
FT DOMAIN 529 565
FT DISULFID 231 240
FT DISULFID 235 246
FT DISULFID 248 257
FT DISULFID 260 271
FT DISULFID 266 277
FT DISULFID 279 288
FT DISULFID 295 307
FT DISULFID 301 317
FT DISULFID 319 328

FT DISULFID 335 BY SIMILARITY. 348
 FT DISULFID 342 BY SIMILARITY. 360
 FT DISULFID 362 BY SIMILARITY. 371
 FT DISULFID 378 BY SIMILARITY. 388
 FT DISULFID 383 BY SIMILARITY. 404
 FT DISULFID 406 BY SIMILARITY. 415
 FT DISULFID 422 BY SIMILARITY. 433
 FT DISULFID 437 BY SIMILARITY. 439
 FT DISULFID 441 BY SIMILARITY. 450
 FT DISULFID 457 BY SIMILARITY. 468
 FT DISULFID 462 BY SIMILARITY. 477
 FT DISULFID 479 BY SIMILARITY. 488
 FT DISULFID 495 BY SIMILARITY. 506
 FT DISULFID 500 BY SIMILARITY. 515
 FT DISULFID 517 BY SIMILARITY. 526
 FT DISULFID 533 BY SIMILARITY. 544
 FT DISULFID 538 BY SIMILARITY. 553
 FT DISULFID 555 BY SIMILARITY. 564
 FT CARBOHYD 98 N-LINKED (GLCNAC. . .) (POTENTIAL). 98
 FT CARBOHYD 137 N-LINKED (GLCNAC. . .) (POTENTIAL). 137
 FT CARBOHYD 167 N-LINKED (GLCNAC. . .) (POTENTIAL). 167
 FT CONFLICT 5 K -> N (IN REF. 1). 5
 FT CONFLICT 67 V -> L (IN REF. 1). 67
 FT CONFLICT 363 A -> R (IN REF. 1). 363
 FT CONFLICT 437 GK -> ET (IN REF. 3). 437
 FT CONFLICT 443 A -> S (IN REF. 3). 443
 FT CONFLICT 459 G -> A (IN REF. 3). 459
 FT CONFLICT 490 S -> T (IN REF. 3). 490
 FT CONFLICT 591 T -> A (IN REF. 1 AND 3). 591
 FT CONFLICT 631 D -> N (IN REF. 1). 631
 FT CONFLICT 652 G -> A (IN REF. 1). 652
 FT CONFLICT 662 L -> M (IN REF. 1). 662

Query Match 80.4%; Score 45; DB 1; Length 833;
 Best Local Similarity 70.0%; Pred. No. 1.8;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGSGDGC 10
 Db 319 CAPGSGDGC 328

RESULT 10
 PCGB_RAT STANDARD; PRT; 883 AA.
 AC P55066; Q63040; Q62860; Q63513;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Brevican core protein precursor (Brain enriched hyaluronan binding protein) (BEHAB protein).
 DE BCAN.
 GN Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Brain;
 RX MEDLINE=96070828; PubMed=752978;
 RA Seidenbecher C.I., Richter K., Rauch U., Faessler R., Garner C.C., Gundelfinger E.D.;
 RA "Brevican, a chondroitin sulfate proteoglycan of rat brain, occurs as secreted and cell surface glycosylphosphatidylinositol-anchored isoforms.";
 RT J. Biol. Chem. 270:27206-27212(1995).
 RL [2]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 396-407.
 RP TISSUE-Brain;
 RC MEDLINE=96074575; PubMed=7488217;
 RX Yamada H., Watanabe K., Shimonaka M., Yamasaki M., Yamauchi Y.;
 RA "cDNA cloning and the identification of an aggrecanase-like cleavage site in rat brevican.";
 RT J. Biol. Chem. 270:27206-27212(1995).

RL Biochem. Biophys. Res. Commun. 216:957-963(1995).
 RN [3]
 RP SEQUENCE OF 1-423 FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Brain;
 RX MEDLINE=94216386; PubMed=7512973;
 RA Jaworski D.M., Kelly G.M., Hockfield S.;
 RT "BEHAB, a new member of the proteoglycan tandem repeat family of hyaluronan-binding proteins that is restricted to the brain.";
 RL J. Cell Biol. 125:455-509(1994).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE TERMINALLY DIFFERENTIATING AND THE ADULT NERVOUS SYSTEM DURING POSTNATAL DEVELOPMENT. COULD STABILIZE INTERACTIONS BETWEEN HA AND BRAIN PROTEOGLYCAN. THE GPI-ANCHORED ISOFORM MAY FUNCTION AS A CHONDROITIN SULFATE-BEARING CELL SURFACE RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX AND ONE FORM ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SECRETED FORM (SHOWN HERE) AND A GPI-ANCHORED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: BRAIN.
 CC -!- DEVELOPMENTAL STAGE: SOLUBLE FORM INCREASES FROM DAY P4 TO P64. GPI-ANCHORED ISOFORM INCREASES AFTER DAY P8.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE.
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 364 ONWARD AND IS SMALLER (371 AA) DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X79881; CAA56255.1; -;
 DR EMBL; X86406; CAA60160.1; -;
 DR EMBL; U37142; AAB87847.1; -;
 DR EMBL; Z28366; CAA82215.1; ALT_FRAME.
 DR HSSP; P20693; 1HLJ.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF 2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00406; IGV; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS01241; LINK; 2.
 DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Lectin; Signal; Sushi.
 KW EGF-like domain; Repeat; Immunoglobulin domain; Alternative splicing; GPI-anchor.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 883 BREVICAN CORE PROTEIN.

```

FT DOMAIN 32 157 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 173 250 LINK 1.
FT DOMAIN 271 352 LINK 2.
FT DOMAIN 622 658 EGF-LIKE.
FT DOMAIN 658 786 C-TYPE LECTIN.
FT DOMAIN 787 851 SUSHI.
FT DISULFID 56 136 BY SIMILARITY.
FT DISULFID 178 249 BY SIMILARITY.
FT DISULFID 202 223 BY SIMILARITY.
FT DISULFID 276 351 BY SIMILARITY.
FT DISULFID 300 321 BY SIMILARITY.
FT DISULFID 626 637 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 648 657 BY SIMILARITY.
FT DISULFID 791 834 BY SIMILARITY.
FT DISULFID 820 847 BY SIMILARITY.
FT CARBOHYD 129 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 625 645 DCIPSPCHNGGTCLEKEGR -> NSAAGSMPAFLFLLL
QWDT (IN GPI-ANCHORED ISOFORM).
FT VARSPLIC 646 883 MISSING (IN GPI-ANCHORED ISOFORM).
FT CONFLICT 51 52 AL -> WV (IN REF. 3).
FT CONFLICT 503 503 V -> L (IN REF. 2).
FT CONFLICT 518 519 TV -> PA (IN REF. 2).
FT CONFLICT 526 526 G -> R (IN REF. 2).
FT CONFLICT 541 541 G -> A (IN REF. 2).
FT CONFLICT 556 556 R -> S (IN REF. 2).
FT CONFLICT 573 573 E -> A (IN REF. 2).
FT CONFLICT 583 583 V -> L (IN REF. 2).
FT CONFLICT 649 649 V -> L (IN REF. 2).
FT CONFLICT 670 670 P -> A (IN REF. 2).
FT CONFLICT 738 738 P -> A (IN REF. 2).
FT CONFLICT 809 809 R -> A (IN REF. 2).
SQ SEQUENCE 883 AA; 96057 MW; AC7ACC40CB53ED37 CRC64;

Query Match 80.4%; Score 45; DB 1; Length 883;
Best Local Similarity 70.0%; Pred. No. 1.9;
Matches 7; Conservations 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVVIGSGDXC 10
DB 648 CVPVGGDLG 657

RESULT 11
EGF_HUMAN STANDARD; PRT; 1207 AA.
AC P01133;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal
DE growth factor (Urogastrone)].
GN EGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=87066721; PubMed=3491360;
RA Bell G.I., Fong N.M., Stempflen M.M., Wormsted M.A., Caput D.,
RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;
RT "Human epidermal growth factor precursor: cDNA sequence, expression
RT in vitro and gene organization.";
RL Nucleic Acids Res. 14:8427-8446(1986).
[2]
RN SEQUENCE OF 971-1023.
RX MEDLINE=77117897; PubMed=300079;
RA Gregory H., Preston B.M.;
RT "The primary structure of human urogastrone.";
RL Int. J. Pept. Protein Res. 9:107-118(1977).

```

```

RN [3]
RP SEQUENCE OF 971-1023.
RX MEDLINE=89391964; PubMed=2789514;
RA Furuya M., Akashi S., Hirayama K.;
RT "The primary structure of human EGF produced by genetic engineering,
RL studied by high-performance tandem mass spectrometry.";
RL Biochem. Biophys. Res. Commun. 163:1100-1106(1989).
[4]
RN STRUCTURE BY NMR OF EGF.
RX MEDLINE=92395667; PubMed=1522591;
RA Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;
RT "Human epidermal growth factor. High resolution solution structure
RT and comparison with human transforming growth factor alpha.";
RL J. Mol. Biol. 227:271-282(1992).
CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X04571; CAA28240.1;
CC PIR: A25531; EGHU.
CC HSSP: P01132; LEGF.
CC Genew: HGNC:3229; EGF.
CC MIM: 131530;
CC
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001336; EGF_1.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR000033; Ldl_receptor_rep.
CC Pfam: PF00008; EGF; 9.
CC Pfam: PF00058; ldl_recept_b; 7.
CC PRINTS: PR00009; EGFTGF.
CC SMART: SM00179; EGF_CA; 2.
CC SMART: SM00001; EGF_like; 7.
CC SMART: SM00135; LY; 8.
CC SMART: SM00010; ASX_HYDROXYL; 3.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01186; EGF_2; 7.
CC PROSITE: PS01187; EGF_CA; 3.
CC EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
KW Signal; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1207 PRO-EPIDERMAL GROWTH FACTOR.
FT CHAIN 971 1023 EPIDERMAL GROWTH FACTOR.
FT DOMAIN 23 1032 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1033 1053 POTENTIAL.
FT DOMAIN 1054 1207 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 314 355 EGF-LIKE 1.
FT DOMAIN 356 396 EGF-LIKE 2.
FT DOMAIN 397 437 EGF-LIKE 3.
FT DOMAIN 435 477 EGF-LIKE 4.
FT DOMAIN 741 781 EGF-LIKE 5.
FT DOMAIN 831 869 EGF-LIKE 6.
FT DOMAIN 870 911 EGF-LIKE 7.
FT DOMAIN 912 952 EGF-LIKE 8.
FT DOMAIN 972 1013 EGF-LIKE 9.
FT DISULFID 318 330 BY SIMILARITY.
FT DISULFID 325 339 BY SIMILARITY.
FT DISULFID 341 354 BY SIMILARITY.
FT DISULFID 360 371 BY SIMILARITY.
FT DISULFID 367 380 BY SIMILARITY.
FT DISULFID 382 395 BY SIMILARITY.
FT DISULFID 401 412 BY SIMILARITY.
FT DISULFID 408 421 BY SIMILARITY.

```

```

FT DISULFID 423 436 BY SIMILARITY.
FT DISULFID 439 451 BY SIMILARITY.
FT DISULFID 447 461 BY SIMILARITY.
FT DISULFID 463 476 BY SIMILARITY.
FT DISULFID 745 756 BY SIMILARITY.
FT DISULFID 752 765 BY SIMILARITY.
FT DISULFID 767 780 BY SIMILARITY.
FT DISULFID 835 846 BY SIMILARITY.
FT DISULFID 840 855 BY SIMILARITY.
FT DISULFID 857 868 BY SIMILARITY.
FT DISULFID 874 888 BY SIMILARITY.
FT DISULFID 881 897 BY SIMILARITY.
FT DISULFID 899 910 BY SIMILARITY.
FT DISULFID 916 929 BY SIMILARITY.
FT DISULFID 923 938 BY SIMILARITY.
FT DISULFID 940 951 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 984 1001 BY SIMILARITY.
FT DISULFID 1003 1012 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 708 708 I -> M.
FT SQ SEQUENCE 1207 AA; 133945 MW; D627DC828EF782E9 CRC64;
Query Match 80.4%; Score 45; DB 1; Length 1207;
Best Local Similarity 60.0%; Pred. NO. 2.6;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVGVSGDXC 10
DB 1003 CVGVYIGERC 1012
||| |::|
||| |::|

RESULT 12
NTC3_RAT
ID NTC3_RAT STANDARD; PRT; 2319 AA.
AC Q9RI72;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
CN NOTCH3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN [2]
RN [3]
RP SEQUENCE FROM N.A.
RA Haritunians T., Boulter J., Weinmaster G., Schanen N.C.;
RA "Rattus norvegicus mRNA for Notch 3."
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN [3]
RX MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT neural progenitor cells to an astroglial fate."
RL Neuron 29:45-55(2001).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RA "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT functional roles for the Notch-DSL signaling system during brain

```

```

RT development."
RL J. Comp. Neurol. 436:167-181(2001).
CC -|- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC -|- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -|- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
CC (CNS) germinal zones and, in early postnatal life, within
CC numerous cells throughout the CNS. It is more highly localized
CC to ventricular germinal zones.
CC -|- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -|- PTM: Phosphorylated (By similarity).
CC -|- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -|- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -|- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF164486; AAD46653.2;
DR HSP: P00740; 1EDM.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00008; EGF; 34.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00010; EGFBL00D.
DR PRINTS: PR00011; EGFAMININ.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00179; EGF_CA; 20.
DR SMART: SM00001; EGF_like; 12.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS02997; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 18.
DR PROSITE: PS00022; EGF_1; 33.
DR PROSITE: PS01186; EGF_2; 26.
DR PROSITE: PS01187; EGF_CA; 17.
KW Receptor; Transcription regulation; Activator; Differentiation;

```


FT DOMAIN 936 1064
DISULFID 23 34 AVIDIN-LIKE.
FT DISULFID 28 43 BY SIMILARITY.
FT DISULFID 45 54 BY SIMILARITY.
FT DISULFID 180 191 BY SIMILARITY.
FT DISULFID 185 200 BY SIMILARITY.
FT DISULFID 202 211 BY SIMILARITY.
FT DISULFID 218 229 BY SIMILARITY.
FT DISULFID 223 238 BY SIMILARITY.
FT DISULFID 240 249 BY SIMILARITY.
FT DISULFID 256 267 BY SIMILARITY.
FT DISULFID 261 276 BY SIMILARITY.
FT DISULFID 278 287 BY SIMILARITY.
FT DISULFID 294 305 BY SIMILARITY.
FT DISULFID 299 314 BY SIMILARITY.
FT DISULFID 316 325 BY SIMILARITY.
FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 337 352 BY SIMILARITY.
FT DISULFID 354 363 BY SIMILARITY.
FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 603 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 750 761 BY SIMILARITY.
FT DISULFID 755 770 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 793 808 BY SIMILARITY.
FT DISULFID 810 819 BY SIMILARITY.
FT DISULFID 826 837 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 848 857 BY SIMILARITY.
FT DISULFID 864 875 BY SIMILARITY.
FT DISULFID 869 884 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 902 913 BY SIMILARITY.
FT DISULFID 907 922 BY SIMILARITY.
FT DISULFID 924 933 BY SIMILARITY.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 477 780 MISSING (IN ISOFORM IB).
FT CONFLICT 279 279 L -> S (IN REF. 2)
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;

Query Match 78.6%; Score 44; DB 1; Length 1064;
Best Local Similarity 60.0%; Pred. No. 3.4;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Oy 1 CVIGYSGDXC 10
Db 316 CPLGFGSDNC 325
Search completed: July 2, 2003, 07:31:33
Job time : 11.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 41.3333 Seconds
(without alignments)
49.850 Million cell updates/sec

Title: US-09-673-785D-9
Perfect score: 56
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	1200	11	Q8VD07 mus musculus
2	51	91.1	2146	5	Q9VC97 drosophila
3	51	91.1	4006	11	O35452 mus musculus
4	51	91.1	4114	11	O54796 mus musculus
5	49	87.5	372	5	Q21756 caenorhabdi
6	48	85.7	191	5	Q8T521 caenorhabdi
7	48	85.7	191	5	Q8T520 caenorhabdi
8	48	85.7	193	5	Q8T522 caenorhabdi
9	48	85.7	193	5	Q8STG0 caenorhabdi
10	48	85.7	194	5	Q8T523 caenorhabdi
11	48	85.7	963	5	Q9GPM9 caenorhabdi
12	48	85.7	1162	5	Q8WTP0 halocynthia
13	48	85.7	1270	5	Q9GPN0 caenorhabdi
14	48	85.7	1531	11	Q9WVB5 mus musculus
15	48	85.7	2920	11	Q9R0M0 mus musculus
16	46	82.1	1530	11	Q9WUG5 rattus norv

17	46	82.1	1531	11	O88279
18	45	80.4	89	6	Q28867
19	45	80.4	403	5	O18375
20	45	80.4	559	5	Q9VZ44
21	45	80.4	775	5	Q8T314
22	45	80.4	1722	5	O19350
23	44	78.6	165	13	Q90XX0
24	44	78.6	264	5	Q20043
25	44	78.6	2352	5	O61240
26	44	78.6	4135	6	O18977
27	44	78.6	4288	4	Q9NPK9
28	43	76.8	225	5	O8T2M9
29	43	76.8	399	5	Q20459
30	43	76.8	464	5	Q95RM9
31	43	76.8	615	13	O57409
32	43	76.8	664	4	Q9UIL7
33	43	76.8	752	13	O42374
34	43	76.8	969	4	Q96KG6
35	43	76.8	1515	13	Q9DE37
36	43	76.8	1534	4	O75093
37	43	76.8	2656	5	Q9GNU3
38	43	76.8	3704	5	P91904
39	42	75.0	106	4	Q9UN93
40	42	75.0	308	4	Q9UN94
41	42	75.0	330	4	Q9NT67
42	42	75.0	366	4	Q9UN95
43	42	75.0	522	4	O8TEP7
44	42	75.0	548	5	O9XUC0
45	42	75.0	790	5	Q9GSF3

ALIGNMENTS

RESULT 1

Q8VD07 PRELIMINARY; PRT; 1200 AA.
ID Q8VD07;
AC Q8VD07;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Similar to epidermal growth factor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017681; AAH17681.1; -;
DR InterPro; IPR000152; ASX_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001336; EGF_1.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 8.
DR Pfam; PF00058; ldl_recept_b; 7.
DR PRINTS; PR00009; EGF_TGF.
DR SMART; SM00181; EGF; 9.
DR SMART; SM00179; EGF_Ca; 8.
DR SMART; SM00135; LY; 9.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_6.
DR PROSITE; PS01187; EGF_Ca; UNKNOWN_3.
SQ SEQUENCE 1200 AA; 131317 MW; B6A44F7294746476 CRC64;

Query Match 100.0%; Score 56; DB 11; Length 1200;
Best Local Similarity 90.0%; Pred. No. 0.12;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CVIGSGDXC 10
Db 992 CVIGSGDRC 1001

RESULT 2
Q9VC97 09VC97 PRELIMINARY; PRT; 2146 AA.
AC Q9VC97;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE CRB protein.
GN CRB OR CG6383.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Miklos G.L.G.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu K., Baldwin D.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamas I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003747; AAF56276.1; -
DR HSSP; P00740; 1EDM.
DR FlyBase; FBgn000368; crb.
DR InterPro; IPR000152; Asx_Hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 26.
DR Pfam; PF00054; Laminin_G; 3.

Query Match 91.1%; Score 51; DB 11; Length 4006;
Best Local Similarity 70.0%; Pred. No. 3.4;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
Db 532 CAVIGSGDRC 541

RESULT 4
Q54796 054796 PRELIMINARY; PRT; 4114 AA.
AC Q54796;
DT 01-JUN-1998 (TREMELrel. 06, Created)

```

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Murray J., Le T.T.;
RT "The sequence of C. elegans cosmid R05G6.";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U98746; AAB00626.1; -.
DR HSSP; P01132; IA3P.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 5.
DR SMART; SM00181; EGF; 7.
DR PROSITE; PS00022; EGF_1; UNKNOWN_7.
DR PROSITE; PS01186; EGF_2; 6.
DR EGF-like domain; Glycoprotein; Hypothetical protein.
SQ EGF-like domain; Glycoprotein; DB36AB251EEB6884 CRC64;
KW SEQUENCE 372 AA; 39085 MW; 39085 MW; 39085 MW;

Query Match 87.5%; Score 49; DB 5; Length 372;
Best Local Similarity 70.0%; Pred. No. 0.67;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDNC 10
I :|||||
DB 233 CYLGYSGDKC 242

RESULT 6
Q8T521 PRELIMINARY; PRT; 191 AA.
ID Q8T521 PRELIMINARY; PRT; 191 AA.
AC Q8T521
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Telodrilinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN-PB800;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491462; AAM09702.1; -.
FT NON_TER 1
FT NON_TER 191
FT NON_TER 191
SQ SEQUENCE 191 AA; 20977 MW; 20977 MW; 20977 MW;

Query Match 85.7%; Score 48; DB 5; Length 191;
Best Local Similarity 70.0%; Pred. No. 0.52;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDNC 10
I :|||||
DB 53 CPLGYSGDNC 62

RESULT 7
Q8T520 PRELIMINARY; PRT; 191 AA.
ID Q8T520 PRELIMINARY; PRT; 191 AA.
AC Q8T520;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VT847;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491463; AAM09703.1; -.
FT NON_TER 1 191
SQ SEQUENCE 191 AA; 20943 MW; 2EEB21B5F47FB470 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 191;
Best Local Similarity 70.0%; Pred. No. 0.52;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10
Db 53 CPLGYSGDYC 62

RESULT 8
O8T522 PRELIMINARY; PRT; 193 AA.
ID O8T522
AC O8T522;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF16;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491460; AAM09700.1; -.
FT NON_TER 1 193
SQ SEQUENCE 193 AA; 21178 MW; 42135BB8B8BE02C5 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 193;
Best Local Similarity 70.0%; Pred. No. 0.52;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10
Db 55 CPLGYSGDYC 64

RESULT 9
O8STG0 PRELIMINARY; PRT; 193 AA.
ID O8STG0;
AC O8STG0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK104, AND HK105;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode

```

```

RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491459; AAM09699.1; -.
DR EMBL; AF491461; AAM09701.1; -.
FT NON_TER 1 193
SQ SEQUENCE 193 AA; 21180 MW; 413C8AB647B5C540 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 193;
Best Local Similarity 70.0%; Pred. No. 0.52;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10
Db 56 CPLGYSGDYC 65

RESULT 10
O8T523 PRELIMINARY; PRT; 194 AA.
ID O8T523
AC O8T523;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB826;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491458; AAM09698.1; -.
FT NON_TER 1 194
SQ SEQUENCE 194 AA; 21275 MW; 07313C8ABD9C65C5 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 194;
Best Local Similarity 70.0%; Pred. No. 0.52;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 CVIGYSGDXC 10
Db 56 CPLGYSGDYC 65

RESULT 11
O9GPM9 PRELIMINARY; PRT; 963 AA.
ID O9GPM9;
AC O9GPM9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Notch-like transmembrane receptor (Fragment).
OS GLP-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF16;
RX MEDLINE=21100341; PubMed=11156985;
RA Rudel D., Kimble J.;
RT "Conservation of glp-1 regulation and function in nematodes.";
RL Genetics 157:639-654(2001).
DR EMBL; AF315555; AAG49317.1; -.
DR HSSP; P01132; IEFG.
DR InterPro; IPR002110; ANK.

```



```

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 4.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00011; EGFLAMININ.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00181; EGF; 5.
DR SMART: SM00001; EGF-like; 4.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS00088; ANK_REPEAT; 3.
DR PROSITE: PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_4.
DR PROSITE: PS01186; EGF_2; 3.
KW ANK repeat; Receptor; Repeat; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 963 AA; 106394 MW; 52BF80010E12FB57 CRC64;

Query Match 85.7%; Score 48; DB 5; Length 963;
Best Local Similarity 70.0%; Pred. No. 2.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDNC 10
| :||||:|
DB 123 CPLYSGDYC 132

RESULT 12
Q8WTP0 PRELIMINARY; PRT; 1162 AA.
AC Q8WTP0;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Vitelline coat component HrvC120.
GN HrvC120.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyuridae; Halocynthia.
OX NCBI_TaxID=7729;
RN [1]
RP SEQUENCE FROM N.A.
RA Sawada H., Sakai N., Abe Y., Tanaka E., Takahashi Y., Fujino J.,
RA Kodama E., Takizawa S., Yokosawa H.;
RT 'A Novel Sperm Receptor VC70, Consisting of 12 EGF-like Repeats, Is
RT Degraded by the Ubiquitin-Proteasome Pathway during Ascidian
RT Fertilization.'
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB061740; BAB7021.1; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR001507; Endoglin/CD105.
DR Pfam: PF00008; EGF; 13.
DR PRINTS: PR00010; EGFBLD.
DR SMART: SM00181; EGF; 14.
DR SMART: SM00241; 2P; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_13.
DR PROSITE: PS01186; EGF_2; UNKNOWN_10.
SQ SEQUENCE 1162 AA; 124499 MW; 26F57DE32B1A8E6C CRC64;

Query Match 85.7%; Score 48; DB 5; Length 1162;
Best Local Similarity 70.0%; Pred. No. 3.3;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDNC 10

```

```

DB 650 CHVGSGDHC 659
| :||||:|

RESULT 13
Q9GPN0 PRELIMINARY; PRT; 1270 AA.
AC Q9GPN0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Notch-like transmembrane receptor.
GN GLP-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF16;
RX MEDLINE=21100341; PubMed=11156985;
RA Rudel D., Kimble J.;
RT "Conservation of glp-1 regulation and function in nematodes.";
RL Genetics 157:639-654 (2001).
DR EMBL: AF315554; AAG49316.1; -.
DR HSSP: P01132; 1EGF.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 10.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00011; EGFLAMININ.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00181; EGF; 12.
DR SMART: SM00179; EGF_Ca; 10.
DR SMART: SM00001; EGF-like; 8.
DR SMART: SM00004; NL; 3.
DR PROSITE: PS00088; ANK_REPEAT; 3.
DR PROSITE: PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 2.
DR PROSITE: PS00022; EGF_1; UNKNOWN_10.
DR PROSITE: PS01186; EGF_2; 8.
DR PROSITE: PS01187; EGF_Ca; 1.
DR ANK repeat; Receptor; Repeat; Transmembrane.
KW ANK repeat; Receptor; Repeat; Transmembrane.
SQ SEQUENCE 1270 AA; 138964 MW; A7662EB575A4B61B CRC64;

Query Match 85.7%; Score 48; DB 5; Length 1270;
Best Local Similarity 70.0%; Pred. No. 3.7;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDNC 10
| :||||:|
DB 430 CPLYSGDYC 439

RESULT 14
Q9WVB5 PRELIMINARY; PRT; 1531 AA.
AC Q9WVB5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SLIT1.
GN SLIT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/ICR;
RX MEDLINE=99365246; PubMed=10433822;
RA Yuan W., Zhou L., Chen J.H., Wu J.Y., Rao Y., Ornitz D.M.;
RT "The mouse Slit family: secreted ligands for ROBO expressed in
RL patterns that suggest a role in morphogenesis and axon guidance.";
RL dev. Biol. 212:290-306(1999).
DR EMBL; AF144627; AA044758.1;
DR HSSP; P00743; ICCF.
DR MGD; MGI:1315203; Slit1.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000359; Cys_knot.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003645; FcIn.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_Out.
DR InterPro; IPR003591; LRR_Typ.
DR Pfam; PF00008; EGF_9.
DR Pfam; PF00054; laminin_G; 1.
DR Pfam; PF00560; LRR; 19.
DR Pfam; PF01463; LRRCT; 4.
DR Pfam; PF01462; LRRNT; 4.
DR PRINTS; PR00010; EGFBLDOD.
DR SMART; SM00041; CT; 1.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_like; 7.
DR SMART; SM00274; FcIn; 2.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00370; LRR; 3.
DR SMART; SM00082; LRRCT; 4.
DR SMART; SM00013; LRRNT; 4.
DR SMART; SM00369; LRR_Typ; 10.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 1531 AA; 167546 MW; F7D09AA6693A4F30 CRC64;

Query Match 85.7%; Score 48; DB 11; Length 1531;
Best Local Similarity 60.0%; Pred. No. 4.4;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 1071 CMLGTTGDC 1080

RESULT 15
Q9ROM0 ID Q9ROM0 PRELIMINARY; PRT; 2920 AA.
AC Q9ROM0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Flamingo 1.
GN CELSR2 OR FLAMINGO 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=99418630; PubMed=10490098;
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RT Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of frizzled.";
RL Cell 98:585-595(1999).
CC -1- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.
DR EMBL; AB028499; BAA84070.1;
DR HSSP; P00740; 1EDM.
DR MGD; MGI:1858235; Celsr2.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR002049; Laminin_G.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PKD_cys_rich.
DR InterPro; IPR001368; TNFR_c6.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin_G; 2.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00180; EGF_Lam; 1.
DR SMART; SM00001; EGF_like; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00282; LamG; 2.
DR SMART; SM00208; TNFR; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS00268; CADHERIN_2; 9.
DR PROSITE; PS00022; EGF_1; UNKNOWN_6.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
KW Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein.
SQ SEQUENCE 2920 AA; 317648 MW; 2919558DF467114F CRC64;

Query Match 85.7%; Score 48; DB 11; Length 2920;
Best Local Similarity 70.0%; Pred. No. 8.6;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDXC 10
DB 1820 CVLGYYGDC 1829

```

Search completed: July 2, 2003, 07:28:38
Job time : 42.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 43.3333 Seconds
(without alignments)
30.750 Million cell updates/sec

Title: US-09-673-785D-13
Perfect score: 50
Sequence: 1 CVIGXSGDCX 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	10	21	AAV52143
2	50	100.0	47	23	AAG66047
3	50	100.0	53	5	AAP40315
4	50	100.0	53	10	AAP91658
5	50	100.0	53	11	AAR08007
6	50	100.0	53	16	AAR67275
7	50	100.0	53	18	AAW34466
8	50	100.0	53	18	AAW34467
9	50	100.0	53	19	AAW50134
10	50	100.0	53	19	AAW50139

11	50	100.0	53	19	AAW50140	Epidermal growth f
12	50	100.0	53	22	AAW37612	Human EGF. Homo s
13	50	100.0	53	23	AAE15714	Epidermal growth f
14	50	100.0	54	23	AAU76704	Mouse epidermal gr
15	50	100.0	54	23	AAU76706	Modified mouse epi
16	50	100.0	56	7	AAP60791	Epidermal growth f
17	50	100.0	117	19	AAW68455	Recombinant human
18	50	100.0	376	22	AAU02933	Angiotensin conver
19	50	100.0	376	22	AAU02956	Angiotensin conver
20	50	100.0	444	22	AAU02934	Angiotensin conver
21	50	100.0	576	22	AAU02930	Angiotensin conver
22	50	100.0	576	22	AAU02955	Angiotensin conver
23	50	100.0	1073	22	AAU02953	Angiotensin conver
24	50	100.0	1080	6	AAP50296	Mouse epidermal gr
25	45	90.0	2146	22	ABG62317	Drosophila melanog
26	43	86.0	655	23	AAW52242	Human versican iso
27	43	86.0	741	21	AAW34485	Human cancer assoc
28	43	86.0	1473	23	AAE18208	Human MOL1b protei
29	43	86.0	2409	12	AAE12609	Versican. Homo sa
30	43	86.0	2469	23	AAE18207	Human MOL1a protei
31	43	86.0	2471	20	AAV06816	Human Notch2 (humN
32	42	84.0	45	21	AAW76904	Zebrafish Delta-B
33	41	82.0	14	18	AAW34460	Chimeric epidermal
34	40	80.0	46	23	AAG66045	Mouse NRG-2 EGF-11
35	40	80.0	48	15	AAW57106	N terminal of rat
36	40	80.0	48	20	AAV01790	Rat epithelial gro
37	40	80.0	71	9	AAW81372	Rat epidermal grow
38	40	80.0	73	9	AAW81367	Rat epidermal grow
39	40	80.0	298	23	AAU11636	Human Neuregulin-2
40	40	80.0	308	22	AAW35196	Human transmembran
41	40	80.0	366	22	AAW35195	Human transmembran
42	40	80.0	366	22	AAW35202	Human transmembran
43	40	80.0	601	23	ABW93437	Herbicideally activ
44	40	80.0	754	18	AAW27535	Rat cerebellum der
45	40	80.0	1296	23	ABG66702	Human novel polype

ALIGNMENTS

RESULT 1
AAV52143
ID AAV52143 standard; peptide: 10 AA.
XX AAV52143;
AC AAV52143;
XX 28-JAN-2000 (first entry)

Mouse EGF derived peptide for targetting laminin receptor.

Epidermal growth factor; EGF; laminin receptor; angiogenesis;
medicament; wound healing; retinopathy of immaturity; metastatic cancer;
candida infection; leishmania; trichomonas vaginalis.

Mus sp.
Key Location/Qualifiers
FT Modified-site 1 /note= "Acetyl-Cys (S-ACM)"
FT Modified-site 10 /note= "Cys (S-ACM)-NH2"
XX WO9954356-A1.
XX 28-OCT-1999.
XX 21-APR-1999; 99WO-GB01211.
XX 22-APR-1998; 98GB-0008407.
XX (UYBE-) UNIV QUEENS BELFAST.
XX Nelson J, Walker B, McFerran N, Harriott P;

XX WPI; 2000-013229/01.
 XX New peptide derived from murine epidermal growth factor (mEGF)
 PT Claim 4; Page 28; 35pp; English.
 PS
 XX This is a peptide derived from mouse epidermal growth factor (EGF)
 CC residues 33-42. This peptide is used in the invention to prepare a
 CC composition to target laminin receptors. EGF derived peptides inhibit
 CC blood vessel formation through their antagonism of the high affinity
 CC laminin receptor found on endothelial cells. The peptide is modified
 CC from the natural sequence to prevent protease attack. The peptide is used
 CC in the preparation of a medicament for binding to laminin receptors as an
 CC (ant)agonist. The medicament is also useful for healing endothelial cell
 CC wounds and treating angiogenic diseases, especially retinopathy of
 CC immaturity. Other diseases treated include metastatic cancer,
 CC Candida spp. infection, and parasitic infestations like leishmania and
 CC trichomonas vaginalis. The peptide are anti-angiogenic in human models.
 CC The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,
 CC and prevent tumour cell attachment to basement membranes.
 XX
 SQ Sequence 10 AA;
 Query Match 100.0%; Score 50; DB 21; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.51;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDXC 10
 ||||:||||:
 Db 1 CVIGYSGDRC 10
 ||||:||||:
 RESULT 2
 AAG66047
 ID AAG66047 standard; peptide; 47 AA.
 XX
 AC AAG66047;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Mouse EGF motif sequence.
 XX
 KW ErbB-4; neuregulin-4; NRG-4; pro-NRG-4; neuroprotective; vulnerary;
 KW cerebroprotective; vasotropic; antiparkinsonian; anticonvulsant;
 KW cytotstatic; nootropic; EGF; NRG-2.
 OS Mus musculus.
 OS WO200181540-A2.
 PN
 XX
 PD 01-NOV-2001.
 XX
 XX 20-APR-2001; 2001WO-IL00371.
 XX
 XX 21-APR-2000; 2000US-0553769.
 XX
 XX (YEDA) YEDA RES & DEV CO LTD.
 XX
 XX Harari D, Yarden Y;
 PI
 XX WPI; 2002-041398/05.
 DR
 XX Novel ErbB-4 ligand, referred as neuregulin (NRG)-4 and polynucleotide
 PT sequences encoding NRG-4, useful for upregulating or downregulating
 PT ErbB-4 receptor activity to treat Alzheimer's disease, stroke, gastric
 PT cancer
 XX
 PS Disclosure; Fig 1c; 153pp; English.
 XX
 XX The invention relates to a novel ErbB-4 ligand, neuregulin-4 (NRG-4).
 CC NRG-4 binds to mammalian ErbB-4 receptor and can be expressed by standard
 CC recombinant methodology. Pharmaceutical compositions comprising NRG-4 are

CC useful for regulating an endogenous protein affecting ErbB-4 receptor
 CC activity in vivo. They are also useful for treating or preventing a
 CC disease condition or syndrome associated with dysregulation of an
 CC endogenous protein affecting ErbB-4 receptor activity, e.g., amyotrophic
 CC lateral sclerosis (Lou Gehrig's disease), Bell's palsy, spinal muscular
 CC atrophy, brain trauma, stroke, ischemia, Alzheimer's disease, Parkinson's
 CC disease, epilepsy, multiple sclerosis, Huntington's chorea, Down's
 CC syndrome, nerve deafness, neuropathy, muscular dystrophy, extramammary
 CC Paget's disease, gastric, pancreatic, prostate, breast and ovarian
 CC cancer, cervical carcinoma, endometrial adenocarcinoma, pancreatic D
 CC cells-somatostatinoma and Zollinger-Ellison syndrome. The agent comprised
 CC in the pharmaceutical composition includes a polypeptide (e.g., a soluble
 CC ligand binding domain of ErbB-4 i.e., IgB4; or a monoclonal, polyclonal,
 CC humanized, single chain antibody or an immunoreactive derivative of an
 CC antibody) capable of binding the endogenous protein affecting ErbB-4
 CC receptor activity. Traceable synthetic/recombinant NRG-4-tagged molecules
 CC can serve as a diagnostic tool in which cells binding NRG-4 can be
 CC measured. Sequences AAG66044-53 represent the EGF-like motifs of various
 CC growth factors.
 XX

SQ Sequence 47 AA;
 Query Match 100.0%; Score 50; DB 23; Length 47;
 Best Local Similarity 80.0%; Pred. No. 2;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 ||||:||||:
 Db 33 CVIGYSGDRC 42
 ||||:||||:
 RESULT 3
 AAP40315
 ID AAP40315 standard; protein; 53 AA.
 XX
 AC AAP40315;
 XX
 DT 22-JUL-1992 (first entry)
 XX
 DE Epidermal growth factor tripentacontapeptide.
 XX
 KW EGF.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Disulfide-bond 6..20
 FT Disulfide-bond 14..31
 FT Disulfide-bond 33..42
 XX
 PN JP59027858-A.
 XX
 XX 14-FEB-1984.
 PD
 XX
 XX 05-AUG-1982; 82JP-0137128.
 PF
 XX
 XX 05-AUG-1982; 82JP-0137128.
 PR
 XX
 XX (NNSH) NIPPON SHINYAKU KK.
 PA
 XX
 XX WPI; 1984-072465/12.
 DR
 XX
 XX Synthesis of epidermal growth factor polypeptide - by condensation
 PT of protected smaller peptide sequences, de:protection then oxidn. to
 PT cyclise.
 XX
 XX Claim1; Page 1; 8pp; Japanese.
 XX
 XX The amino acid sequence is that of an epidermal growth factor
 CC tripentacontapeptide which is synthesised by condensation of
 CC protected smaller peptide sequences. This method produces the
 CC peptide smoothly, with high purity and yield.
 CC

SQ Sequence 53 AA;
 Query Match 100.0%; Score 50; DB 5; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
 III:III:
 DB 33 CVIGSGDNC 42

RESULT 4
 AAP91658
 ID AAP91658 standard; protein; 53 AA.
 XX
 AC AAP91658;
 XX
 DT 29-JUN-1990 (first entry)
 XX
 DE Amino acid sequence for naturally occurring epidermal growth factor
 (EGF).
 DE
 XX
 KW Epidermal growth factor; angiogenesis; synthetic peptide.
 XX
 PN WO8901489-A.
 XX
 XX
 PD 23-FEB-1989.
 XX
 XX
 PF 10-AUG-1988; 88WO-AU00300.
 XX
 PR 10-AUG-1987; 87AU-0003629.
 XX
 XX
 PA (CSIR) COMMONWEALTH SCIENT. ORG.
 XX
 PI McAuslan BR;
 XX
 DR WPI; 1989-068852/09.
 XX
 XX
 PT Synthetic peptide active in stimulating angiogenesis -
 PT has sequences corresponding to amino acid sequences occurring in
 PT epidermal growth factor.
 XX
 XX
 PS Fig 1; 1/1; 1lpp; English.
 XX
 CC The inventors claim synthetic peptides which correspond to sequences
 CC occurring in EGF, but excluding EGF. The synthetic peptides
 CC correspond to active sequences from EGF fragments: 3-14, 12-14, 12-15,
 CC 3-10, 29-37 and 33-37 (Claims 2-7 and AAP91659-p91664 resp.). The
 CC peptides are angiogenic. Their relative shortness means that they pose
 CC fewer synthesis problems than the entire EGF molecule.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 50; DB 10; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
 III:III:
 DB 33 CVIGSGDNC 42

RESULT 5
 AAR08007
 ID AAR08007 standard; protein; 53 AA.
 XX
 AC AAR08007;
 XX
 DT 25-FEB-1991 (first entry)
 XX
 XX Modified murine epidermal growth factor.
 DE
 XX Modified murine epidermal growth factor; stability; storage;
 KW

KW epithelial wounds; gastric acid secretion.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 11.11
 FT /label- E, N, Q, A, K
 XX
 PN WO9013570-A.
 XX
 PD 15-NOV-1990.
 XX
 PF 09-MAY-1990; 90WO-US02600.
 XX
 PR 12-MAY-1989; 89US-0351773.
 XX
 PA (CHIR-) CHIRON CORP.
 XX
 XX Nascimento CG, Medina-Selby A;
 PI WPI; 1990-361427/48.
 DR
 XX Human epidermal growth factor - is substituted at position 11 for
 PT greater stability and improved storage life.
 PT
 XX Claim 9; Page 25; 32pp; English.
 PS
 XX The human EGF is used to treat oversecretion of gastric acid or an
 CC epithelial wound. EGF is modified to increase its chemical
 CC stability. Its storage life is improved without diminishing its
 CC biological activity. The proteins may be prepared by traditional
 CC chemical or recombinant means.
 CC See also AAR08004.
 XX
 SQ Sequence 53 AA;
 Query Match 100.0%; Score 50; DB 11; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
 III:III:
 DB 33 CVIGSGDNC 42

RESULT 6
 AAR67275
 ID AAR67275 standard; peptide; 53 AA.
 XX
 AC AAR67275;
 XX
 DT 01-AUG-1995 (first entry)
 XX
 XX Synthetic human/mouse EGF.
 DE
 XX Cell growth factor; viscoelastic solution; fibroblastic growth factor;
 KW EGF; epidermal growth factor; EGF; buffered solution; lubrication;
 KW carboxypropyl methylcellulose; cellulose gum; dextran; dextran sulphate;
 KW chondroitin sulphate; sodium hyaluronate; osmolality; mitogenic;
 KW wound healing; cell protection; cell coating; surgery; tissue space;
 KW hydroxypropyl methylcellulose; manipulation.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 37
 FT /note= "Given in the specification as Try"
 XX
 PN US5366964-A.
 XX
 PD 22-NOV-1994.
 XX
 XX 15-DEC-1988; 88US-0284533.
 PF
 XX 15-DEC-1988; 88US-0284533.
 PR

PR 13-NOV-1989; 89US-0434305.
 XX (LIND/) LINDSTROM R L.
 PA (SKEL/) SKELNIK D.
 XX
 PI Lindstrom RL, Skelnik D;
 XX WPI; 1995-005865/01.
 DR

XX Viscoelastic soln. contg. hydroxypropyl methylcellulose, sodium
 PT hyaluronate, chondroitin sulphate and growth factors - used as a
 PT surgical soln. to promote wound healing, e.g. of corneal cells
 PS Disclosure; Column 5; 8pp; English.
 XX

CC The sequences given in AAR67273-76 are cell growth factors which may
 CC be used in the viscoelastic solution of the invention. The peptides
 CC are derived from fibroblastic growth factor (FGF) and epidermal
 CC growth factor (EGF). The viscoelastic solution also comprises a
 CC buffered solution which is pref. a buffered balanced salt solution,
 CC at least 1 of hydroxy- or carboxypropyl methylcellulose, a cellulose
 CC gum, dextran or dextran sulphate, chondroitin sulphate, and sodium
 CC hyaluronate. It has a pH of 6-8 and an osmolarity of 200-400 mOsmol/l.
 CC The growth factors are mitogenic in vitro for a wide range of tissues
 CC and the viscoelastic soln. may be used as a surgical soln. which is
 CC in direct contact with cells undergoing wound healing. It also provides
 CC a cell protection and cell coating during surgery. The soln. provides
 CC maintenance of tissue space, hydroxypropyl methylcellulose and
 CC chondroitin sulphate lubricate the tissue, while sodium hyaluronate
 CC provides tissue manipulation.
 XX

SQ Sequence 53 AA;

Query Match 100.0%; Score 50; DB 16; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 ||||:||||:
 DB 33 CVIGYSGDRC 42

RESULT 7
 AAW34466
 ID AAW34466 standard; Protein; 53 AA.
 XX
 AC AAW34466;
 XX

DT 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein AC.

DE Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

KW Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

Key	Location/Qualifiers
FT Region	1..10
FT	/note= "region A from mouse EGF"
FT Region	11..33
FT	/note= "region B from human EGF"
FT Region	34..47
FT	/note= "region C from mouse EGF"
FT Region	48..53
FT	/note= "region D from human EGF"

XX AU9717759-A.

PN 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.
 PF

XX

PR 23-APR-1996; 96JP-0123970.
 XX (HGET) HIGETA SHOYU KK.
 PA
 XX Tagami H, Tanaka A;
 PI
 XX WPI; 1997-550187/51.
 DR N-PSDB; AAT99941.

XX Chimeric epidermal growth factor proteins - and DNA molecules for
 PT their recombinant production
 PT
 XX Claim 5; Page 26; 40pp; English.
 PS

XX This is a chimeric epidermal growth factor (EGF) protein AC. This
 CC chimeric EGF protein contains sequences derived from human and mouse
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced
 CC by culturing a microorganism of the genus *Bacillus* transformed by an
 CC expression vector containing the encoding DNA molecules. The recombinant
 CC chimeric EGF proteins exhibit a wide variety of physiological activities
 CC similar to those of naturally occurring EGF.
 XX

SQ Sequence 53 AA;

Query Match 100.0%; Score 50; DB 18; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 ||||:||||:
 DB 33 CVIGYSGDRC 42

RESULT 8
 AAW34467
 ID AAW34467 standard; Protein; 53 AA.
 XX
 AC AAW34467;
 XX

DT 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein ABC.

DE Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

KW Chimeric - Mus sp.

XX Chimeric - Homo sapiens.

Key	Location/Qualifiers
FT Region	1..10
FT	/note= "region A from mouse EGF"
FT Region	11..33
FT	/note= "region B from mouse EGF"
FT Region	34..47
FT	/note= "region C from mouse EGF"
FT Region	48..53
FT	/note= "region D from human EGF"

XX AU9717759-A.

PN 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.
 PF

XX 23-APR-1996; 96JP-0123970.
 PR

XX (HGET) HIGETA SHOYU KK.
 PA

XX Tagami H, Tanaka A;
 PI

XX WPI; 1997-550187/51.
 DR

XX N-PSDB; AAT99942.
 DR

us-09-673-785d-13.jull.rag

Wed Jul 2, 08:08:16 2003

PT Chimeric epidermal growth factor proteins - and DNA molecules for
 their recombinant production

XX Claim 6; Page 26; 40pp; English.

PS This is a chimeric epidermal growth factor (EGF) protein ABC. This
 CC chimeric EGF protein contains sequences derived from human and mouse
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced
 CC by culturing a microorganism of the genus *Bacillus* transformed by an
 CC expression vector containing the encoding DNA molecules. The recombinant
 CC chimeric EGF proteins exhibit a wide variety of physiological activities
 CC similar to those of naturally occurring EGF.

XX Sequence 53 AA;

Query Match 100.0%; Score 50; DB 18; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 Db 33 CVIGYSGDRC 42

RESULT 9

AAW50134
 ID AAW50134 standard; protein; 53 AA.

XX AC AAW50134;

XX 08-JUL-1998 (first entry)

XX Mouse epidermal growth factor.

DE Mouse; epidermal growth factor; EGF; isolation; recombinant;
 KW ion exchange chromatography; *Bacillus brevis*.

XX Mus sp.

XX AU9728698-A.

XX 05-FEB-1998.

XX 17-JUL-1997; 97AU-0028698.

XX 01-AUG-1996; 96JP-0218109.

XX (HGET) HIGETA SHOYU KK.

XX Miyauchi A, Nemoto A, Warren B;

XX WPI; 1998-131057/13.

XX Isolation of recombinant epidermal growth factor from whole broth -
 PT by fluidised-bed ion-exchange chromatography

XX Claim 6; Page 26; 38pp; English.

XX The present sequence represents mouse epidermal growth factor (EGF)
 CC from the present invention. The present invention describes a method for
 CC the isolation of recombinant epidermal growth factor (EGF) from whole
 CC broth. The method comprises: (a) passing a culture containing the
 CC protein upwards through a column containing a fluidised bed of ion
 CC exchanger, washing the ion exchanger to remove non-adsorbed material,
 CC allowing an eluant downwards through the column; (b) filtering the eluate
 CC through a membrane with a molecular weight cutoff of 10000, and (c)
 CC concentrating the filtrate on a membrane with a molecular weight cutoff
 CC of 5000. The method is for isolating recombinant human, mouse, pig or
 CC rat EGF, or an EGF analogue with one of four defined amino acid sequences
 CC (given in the specification), from *Bacillus brevis* cultures. The EGF
 CC proteins can be recovered in high yield and high purity without the need
 CC to pretreat the culture broth to remove the cells, e.g. by costly

CC centrifugation or membrane filtration.

XX Sequence 53 AA;

Query Match 100.0%; Score 50; DB 19; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 Db 33 CVIGYSGDRC 42

RESULT 10

AAW50139
 ID AAW50139 standard; protein; 53 AA.

XX AC AAW50139;

XX 08-JUL-1998 (first entry)

XX Epidermal growth factor analogue 3.

XX Epidermal growth factor; EGF; isolation; recombinant;
 KW ion exchange chromatography; *Bacillus brevis*.

XX Unidentified.

XX AU9728698-A.

XX 05-FEB-1998.

XX 17-JUL-1997; 97AU-0028698.

XX 01-AUG-1996; 96JP-0218109.

XX (HGET) HIGETA SHOYU KK.

XX Miyauchi A, Nemoto A, Warren B;

XX WPI; 1998-131057/13.

XX Isolation of recombinant epidermal growth factor from whole broth -
 PT by fluidised-bed ion-exchange chromatography

XX Claim 6; Page 28; 38pp; English.

XX The present sequence represents a protein which has epidermal growth
 CC factor (EGF) like activity, from the present invention. The present
 CC invention describes a method for the isolation of recombinant epidermal
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing
 CC a culture containing the protein upwards through a column containing a
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-
 CC adsorbed material, allowing the ion exchanger to settle, and eluting
 CC adsorbed protein by passing an eluant downwards through the column; (b)
 CC filtering the eluate through a membrane with a molecular weight cutoff
 CC of 10000, and (c) concentrating the filtrate on a membrane with a
 CC molecular weight cutoff of 5000. The method is for isolating recombinant
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four
 CC defined amino acid sequences (given in the specification), from *Bacillus*
 CC *brevis* cultures. The EGF proteins can be recovered in high yield and
 CC high purity without the need to pretreat the culture broth to remove the
 CC cells, e.g. by costly centrifugation or membrane filtration.

XX Sequence 53 AA;

Query Match 100.0%; Score 50; DB 19; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 Db 33 CVIGYSGDRC 42

RESULT 11
AAW50140
ID AAW50140 standard; protein; 53 AA.
XX AC AAW50140;
XX DT 08-JUL-1998 (first entry)
XX DE Epidermal growth factor analogue 4.
XX KW Epidermal growth factor; EGF; isolation; recombinant;
KW ion exchange chromatography; Bacillus brevis.
XX OS Unidentified.
XX PN AU9728698-A.
XX PD 05-FEB-1998.
XX PF 17-JUL-1997; 97AU-0028698.
XX PR 01-AUG-1996; 96JP-0218109.
XX PA (HGET.) HIGETA SHOYU KK.
XX PI Miyauchi A, Nemoto A, Warren B;
XX WPI; 1998-131057/13.
XX PT Isolation of recombinant epidermal growth factor from whole broth -
PT by fluidised-bed ion-exchange chromatography
XX PS Claim 6; Page 29; 38pp; English.
XX CC The present sequence represents a protein which has epidermal growth
CC factor (EGF) like activity, from the present invention. The present
CC invention describes a method for the isolation of recombinant epidermal
CC growth factor (EGF) from whole broth. The method comprises: (a) passing
CC a culture containing the protein upwards through a column containing a
CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-
CC adsorbed material, allowing the ion exchanger to settle, and eluting
CC adsorbed protein by passing an eluant downwards through the column; (b)
CC filtering the eluate through a membrane with a molecular weight cutoff
CC of 10000, and (c) concentrating the filtrate on a membrane with a
CC molecular weight cutoff of 5000. The method is for isolating recombinant
CC human, mouse, pig or rat EGF, or an EGF analogue with one of four
CC defined amino acid sequences (given in the specification), from Bacillus
CC brevis cultures. The EGF proteins can be recovered in high yield and
CC high purity without the need to pretreat the culture broth to remove the
CC cells, e.g. by costly centrifugation or membrane filtration.
XX SQ Sequence 53 AA;
Query Match 100.0%; Score 50; DB 19; Length 53;
Best Local Similarity 80.0%; Pred. No. 2.3;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 CVIGXSGDNC 10
DB 33 CVIGYSGDRC 42
RESULT 12
AAB37612
ID AAB37612 standard; protein; 53 AA.
XX AC AAB37612;
XX DT 27-FEB-2001 (first entry)
XX DE Human EGF.
XX Merozoite surface protein; protazoacide; vaccine; malaria; human; EGF;
XX epidermal growth factor.
XX Homo sapiens.
XX PN WO200063245-A2.
XX PD 26-OCT-2000.
XX PF 20-APR-2000; 2000WO-GB01558.
XX PR 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 990S-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX (MEDI-) MEDICAL RES COUNCIL.
XX PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX WPI; 2001-015762/02.
XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria -
XX PS Disclosure; Fig 1; 126pp; English.
XX CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1₄₋₂, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1₄₋₂, compared to natural MSP-1₁₋₉. The non-natural variants of
CC the present invention are useful for immunising a mammal against malaria,
CC and can be used to treat malaria. The present sequence is human
CC epidermal growth factor (EGF). This sequence was used in a sequence
CC homology comparison with the wild-type MSP-1 protein from P. falciparum
CC (see AAB37608), which was used to generate the variants of the present
CC invention.
XX SQ Sequence 53 AA;
Query Match 100.0%; Score 50; DB 22; Length 53;
Best Local Similarity 80.0%; Pred. No. 2.3;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 CVIGXSGDNC 10
DB 33 CVIGYSGDRC 42
RESULT 13
AAE15714
ID AAE15714 standard; Protein; 53 AA.
XX AC AAE15714;
XX DT 12-MAR-2002 (first entry)
XX DE Epidermal growth factor (EGF) used in exemplification of the invention.
XX KW Humanised form; monoclonal antibody alpha 340; gene therapy;
KW epidermal growth factor receptor; EGF; cancer; colorectal; lung; breast;
KW gastric; ovarian; immune response; cytostatic; cell growth; apoptosis;
XX inhibitor.
XX OS Unidentified.
XX PN WO200188138-A1.
XX PD 22-NOV-2001.

XX PF 21-MAY-2001; 2001WO-GB02226.
 XX PR 19-MAY-2000; 2000GB-0011981.
 XX PR 24-AUG-2000; 2000GB-0020794.
 XX PA (SCAN-) SCANCELL LTD.
 XX XX Ellis JRM, Durrant LG;
 XX WPI; 2002-062384/08.
 XX PR New humanized form of mouse monoclonal antibody 340 which binds to
 XX PT epidermal growth factor receptor and inhibits binding of growth factor,
 XX PT useful for treating colorectal, lung, breast, gastric and ovarian
 XX PT cancer -
 XX PS Example 5; Fig 12; 53pp; English.
 XX XX The present invention relates to a humanised form of the antibody 340 (a
 XX CC mouse monoclonal antibody which binds to epidermal growth factor (EGF)
 XX CC receptor and inhibits binding of EGF), obtainable from the cell line
 XX CC deposited with the ECACC under accession number 97021428. The humanised
 XX CC form of the antibody 340 is useful in gene therapy, medicine and in the
 XX CC manufacture of a medicament for treatment or prophylaxis of cancer. The
 XX CC invention is useful for treating colorectal, lung, breast, gastric or
 XX CC ovarian cancers or also for preventing the recurrence of cancer after
 XX CC initial treatment or surgery. The invention is also useful for enhancing
 XX CC a protective immune response against cancer by optimised immunisation
 XX CC schedules. The humanised form of the antibody 340 has reduced
 XX CC immunogenicity but shows similar binding to cells expressing EGF
 XX CC receptor, as the original murine antibody and has increased ability to
 XX CC inhibit the growth of EGF receptor expressing cells. The invention is
 XX CC used as cell growth and apoptosis inhibitor. The present sequence
 XX CC is epidermal growth factor (EGF) which is used in the exemplification of
 XX CC the invention.
 XX SQ Sequence 53 AA;
 Query Match 100.0%; Score 50; DB 23; Length 53;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDXC 10
 ||||:||||:
 Db 33 CVIGYSGDRC 42
 RESULT 14
 AAU76704
 ID AAU76704 standard; Protein; 54 AA.
 XX AC AAU76704;
 XX XX 21-MAY-2002 (first entry)
 XX DT Mouse epidermal growth factor (EGF).
 XX DE Mouse; epidermal growth factor; receptor; EGF;
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;
 KW recombination.
 XX OS Mus sp.
 XX XX Key Location/Qualifiers
 FT Misc-difference 54
 FT /label= Unknown
 FT /note= "Encoded by TAA"
 XX XX WO200206469-A2.
 XX PN 24-JAN-2002.
 XX PD
 XX XX

PF 18-JUL-2001; 2001WO-US22640.
 XX PR 18-JUL-2000; 2000US-218921P.
 XX PR 18-JUL-2000; 2000US-219085P.
 XX PR 19-OCT-2000; 2000US-0691873.
 XX PR 19-OCT-2000; 2000US-0692732.
 XX PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.
 XX XX COCO WM, Enccell LP, Arensdorf JJ;
 XX WPI; 2002-179795/23.
 XX DR N-PSDB; ABK10313.
 XX XX Forming chimeric polynucleotide by contacting single-stranded (ss)
 XX PT scaffold fragment, and donor fragment populations to form hybridised
 XX PT complex having ss regions which are filled-in, and ligating adjacent
 XX PT fragments -
 XX PS Example 2; Fig 3; 65pp; English.
 XX XX The invention describes a method of forming (M1) chimeric
 XX CC polynucleotides (I). The method comprises contacting a population of
 XX CC single-stranded scaffold fragments (SF) with population of donor
 XX CC fragments (DF) to form a complex (II) comprising at least one SF
 XX CC hybridised to two DFs. (II) is treated so that single stranded regions of
 XX CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for
 XX CC a directed evolution process which involves forming a library of (I)
 XX CC that can be screened for a characteristic of interest. Subsequent rounds
 XX CC of directed evolution can produce chimeric polynucleotides with an
 XX CC improved characteristic of interest. The methods facilitate the
 XX CC generation of chimeric polynucleotides and do not require hybridising
 XX CC donor fragments to a target- or full-length template. Because the
 XX CC chimeraogenesis process does not rely upon a contiguous, full-length
 XX CC template, it is unnecessary to modify a template to facilitate its
 XX CC removal. This is the amino acid sequence of the mouse epidermal growth
 XX CC factor (EGF) used to demonstrate a method of in vitro recombination
 XX CC described in the method of the invention.
 XX SQ Sequence 54 AA;
 Query Match 100.0%; Score 50; DB 23; Length 54;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXSGDXC 10
 ||||:||||:
 Db 33 CVIGYSGDRC 42
 RESULT 15
 AAU76706
 ID AAU76706 standard; Protein; 54 AA.
 XX AC AAU76706;
 XX XX 21-MAY-2002 (first entry)
 XX DT Modified mouse epidermal growth factor (EGF).
 XX DE Mouse; epidermal growth factor; receptor; EGF;
 KW chimeric polynucleotide; directed evolution; chimeraogenesis;
 KW recombination.
 XX OS Mus sp.
 XX XX Key Location/Qualifiers
 FT Misc-difference 38
 FT /note= "Encoded by ACC"
 XX XX WO200206469-A2.
 XX PN
 XX PD
 XX XX

PD 24-JAN-2002.
 XX
 PF 18-JUL-2001; 2001WO-US22640.
 XX
 PR 18-JUL-2000; 2000US-218921P.
 PR 18-JUL-2000; 2000US-219085P.
 PR 19-OCT-2000; 2000US-0691873.
 PR 19-OCT-2000; 2000US-0692732.
 XX
 PA (ENCH-) ENCHIRA BIOTECHNOLOGY CORP.
 XX
 XX Coco WM, Ensell LP, Arensdorf JJ;
 PI WPI: 2002-179795/23.
 XX N-PSDB; ABR10315.
 DR
 DR
 XX
 PT Forming chimeric polynucleotide by contacting single-stranded (ss)
 PT scaffold fragment, and donor fragment populations to form hybridised
 PT complex having ss regions which are filled-in, and ligating adjacent
 PT fragments
 XX
 PS Example 2; Fig 3; 65pp; English.
 XX
 CC The invention describes a method of forming (M1) chimeric
 CC polynucleotides (I). The method comprises contacting a population of
 CC single-stranded scaffold fragments (SF) with population of donor
 CC fragments (DF) to form a complex (II) comprising at least one SF
 CC hybridised to two Dfs. (II) is treated so that single stranded regions of
 CC (II) are filled-in and adjacent fragments are ligated. (M1) is useful for
 CC a directed evolution process which involves forming a library of (I)
 CC that can be screened for a characteristic of interest. Subsequent rounds
 CC of directed evolution can produce chimeric polynucleotides with an
 CC improved characteristic of interest. The methods facilitate the
 CC generation of chimeric polynucleotides and do not require hybridising
 CC donor fragments to a target- or full-length template. Because the
 CC chimeragenesis process does not rely upon a contiguous, full-length
 CC template, it is unnecessary to modify a template to facilitate its
 CC removal. This is the amino acid sequence of a mouse epidermal growth
 CC factor (EGF) modified to make it as similar as possible to the human
 CC protein (AAU76705) and used to demonstrate a method of in vitro
 CC recombination described in the method of the invention.
 XX
 SQ Sequence 54 AA;

Query Match 100.0%; Score 50; DB 23; Length 54;
 Best Local Similarity 80.0%; Pred. No. 2,3;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXGDXC 10
 IIII:III:I
 Db 33 CVIGSGDRC 42

Search completed: July 2, 2003, 07:30:55
 Job time : 43.3333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 10.6667 Seconds
(without alignments)
27.584 Million cell updates/sec

Title: US-09-673-785D-13
Perfect score: 50
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	12	1 US-08-597-545-7	Sequence 7, Appl
2	50	100.0	12	1 US-08-457-135-7	Sequence 7, Appl
3	50	100.0	53	1 US-08-284-923-2	Sequence 2, Appl
4	50	100.0	53	4 US-08-619-032B-2	Sequence 2, Appl
5	50	100.0	53	6 5332669-2	Patent No. 5332669
6	43	86.0	2409	6 5180808-2	Patent No. 5180808
7	43	86.0	2471	1 US-08-185-432-16	Sequence 16, Appl
8	43	86.0	2471	1 US-08-083-590A-19	Sequence 19, Appl
9	43	86.0	2471	3 US-08-532-384-19	Sequence 19, Appl
10	43	86.0	2471	4 US-08-899-232-1	Sequence 1, Appl
11	42	84.0	154	3 US-09-191-647-10	Sequence 10, Appl
12	42	84.0	154	4 US-09-540-245A-10	Sequence 10, Appl
13	42	84.0	154	4 US-09-540-153-10	Sequence 10, Appl
14	40	80.0	48	6 5332669-1	Patent No. 5332669
15	40	80.0	754	2 US-08-525-864A-2	Sequence 2, Appl
16	39	78.0	17	6 5183805-1	Patent No. 5183805
17	39	78.0	17	6 5183805-2	Patent No. 5183805
18	39	78.0	25	6 5256643-8	Patent No. 5256643
19	39	78.0	37	2 US-08-039-364-14	Sequence 14, Appl
20	39	78.0	37	4 US-09-158-710-14	Sequence 14, Appl
21	39	78.0	42	1 US-08-168-091A-44	Sequence 44, Appl
22	39	78.0	44	1 US-08-278-089A-24	Sequence 24, Appl
23	39	78.0	46	3 US-08-899-437-12	Sequence 12, Appl
24	39	78.0	46	4 US-09-126-121-12	Sequence 12, Appl
25	39	78.0	47	3 US-08-753-007A-17	Sequence 17, Appl
26	39	78.0	47	4 US-09-398-496-17	Sequence 17, Appl
27	39	78.0	48	4 US-09-020-880-15	Sequence 15, Appl

28 39 78.0 48 4 US-08-915-096A-13 Sequence 13, Appl
29 39 78.0 48 4 US-09-101-544-15 Sequence 15, Appl
30 39 78.0 48 6 5434135-3 Patent No. 5434135
31 39 78.0 51 6 5177197-50 Patent No. 5177197
32 39 78.0 53 1 US-07-869-176-1 Sequence 1, Appl
33 39 78.0 53 1 US-08-284-923-1 Sequence 1, Appl
34 39 78.0 53 1 US-08-360-841-2 Sequence 2, Appl
35 39 78.0 53 2 US-08-861-000-1 Sequence 1, Appl
36 39 78.0 53 4 US-08-619-032B-1 Sequence 1, Appl
37 39 78.0 53 6 5332669-3 Patent No. 5332669
38 39 78.0 53 6 5434135-2 Patent No. 5434135
39 39 78.0 55 6 5218093-1 Patent No. 5218093
40 39 78.0 88 2 US-07-885-089B-13 Sequence 13, Appl
41 39 78.0 91 1 US-07-847-743B-15 Sequence 15, Appl
42 39 78.0 91 1 US-08-456-201-15 Sequence 15, Appl
43 39 78.0 91 2 US-08-330-161-13 Sequence 13, Appl
44 39 78.0 91 2 US-08-456-241-15 Sequence 15, Appl
45 39 78.0 91 2 US-08-440-401-13 Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-597-545-7
; Sequence 7, Application US/08597545
; Patent No. 5580738
; GENERAL INFORMATION:
; APPLICANT: LABORDA, Jorge
; TITLE OF INVENTION: Delta-Like Gene Expressed In
; TITLE OF INVENTION: Neuroendocrine Tumors
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,545
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/989,537
; FILING DATE: 11-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/166 NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-597-545-7

Query Match 100.0%; Score 50; DB 1; Length 12;
Best Local Similarity 80.0%; pred. No. 0.11;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

Db 3 CVIGSGDRC 12

RESULT 2
US-08-457-135-7
; Sequence 7, Application US/08457135
; Patent No. 5644031
; GENERAL INFORMATION:
; APPLICANT: LABORDA, Jorge
; TITLE OF INVENTION: Delta-Like Gene Expressed In
; TITLE OF INVENTION: Neuroendocrine Tumors
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,135
; FILING DATE: 01-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/989,537
; FILING DATE: 11-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40399/304/NIHD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-457-135-7

Query Match 100.0%; Score 50; DB 1; Length 12;
Best Local Similarity 80.0%; Pred. No. 0.11;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
| | | | | | | |
Db 3 CVIGYSGDRC 12

RESULT 3
US-08-284-923-2
; Sequence 2, Application US/08284923
; Patent No. 5547935
; GENERAL INFORMATION:
; APPLICANT: Mullenbach, Guy T
; APPLICANT: Blaney, Jeffrey M
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: Muteins of Epidermal Growth Factor
; TITLE OF INVENTION: exhibiting enhanced binding at low ph
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,923
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/746,651
; FILING DATE: 16-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcclung, Barbara G
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 231,001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-284-923-2
Query Match 100.0%; Score 50; DB 1; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.44;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
| | | | | | | |
Db 33 CVIGYSGDRC 42

RESULT 4
US-08-619-032B-2
; Sequence 2, Application US/08619032B
; Patent No. 6191106
; GENERAL INFORMATION:
; APPLICANT: Mullenbach, Guy T.
; APPLICANT: Blaney, Jeffrey M.
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: MUTEINS OF EPIDERMAL GROWTH FACTOR
; TITLE OF INVENTION: EXHIBITING ENHANCED BINDING AT LOW PH
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: Intellectual Property R-440, P.O. Box 8097
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,032B
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Guth, Joseph H.
; REGISTRATION NUMBER: 31,261
; REFERENCE/DOCKET NUMBER: 0231,004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3888
; TELEFAX: (510) 655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-619-032B-2

Query Match      100.0%; Score 50; DB 4; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.44;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      ||||:||||-|
Db      33 CVIGYSGDRC 42

RESULT 5
5332669-2
;PATENT NO. 5332669
; APPLICANT: DEUEL, THOMAS F.
; TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/356,739
; FILING DATE: 24-MAR-1989
; SEQ ID NO: 2:
; LENGTH: 53
5332669-2

Query Match      100.0%; Score 50; DB 6; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.44;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      ||||:||||-|
Db      33 CVIGYSGDRC 42

RESULT 6
5180808-2
;PATENT NO. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO: 2:
; LENGTH: 2409
5180808-2

Query Match      86.0%; Score 43; DB 6; Length 2409;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      || |:|||:|
Db      2128 CVPGYSGDQC 2137

RESULT 7
US-08-185-432-16
; Sequence 16, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

;
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741

;
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
;
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-16

Query Match      86.0%; Score 43; DB 1; Length 2471;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVIGXSGDXC 10
      || |:|||:|
Db      326 CVNGWGGDC 335

RESULT 8
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
```

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-083-590A-19

Query Match 86.0%; Score 43; DB 1; Length 2471;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|||:||||:
Db 326 CVNGWSGDDC 335

RESULT 9

US-08-532-384-19
Sequence 19, Application US/08532384
Patent No. 6083904
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-19

Query Match 86.0%; Score 43; DB 3; Length 2471;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|||:||||:
Db 326 CVNGWSGDDC 335

RESULT 10
US-08-899-232-1
Sequence 1, Application US/08899232
Patent No. 6436650
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Qi, Huilin
TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
FILE REFERENCE: 7326-046
CURRENT APPLICATION NUMBER: US/08/899,232
CURRENT FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2471
TYPE: PRT
ORGANISM: Homo sapiens
US-08-899-232-1

Query Match 86.0%; Score 43; DB 4; Length 2471;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|||:||||:
Db 326 CVNGWSGDDC 335

RESULT 11

US-09-191-647-10
Sequence 10, Application US/09191647
Patent No. 6046015
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions
FILE REFERENCE: B98-031-3
CURRENT APPLICATION NUMBER: US/09/191,647
CURRENT FILING DATE: 1998-11-13
EARLIER APPLICATION NUMBER: 60/065,544
EARLIER FILING DATE: 1997-11-14
EARLIER APPLICATION NUMBER: 60/081,057
EARLIER FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 154
TYPE: PRT
ORGANISM: mouse
US-09-191-647-10

Query Match 84.0%; Score 42; DB 3; Length 154;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|||:||||:
Db 11 CMLGYTGDC 20

RESULT 12

US-09-540-245A-10
Sequence 10, Application US/09540245A
Patent No. 6270984
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions

FILE REFERENCE: B98-031-3
CURRENT APPLICATION NUMBER: US/09/540,245A
CURRENT FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 60/065,544
PRIOR FILING DATE: 1997-11-14
PRIOR APPLICATION NUMBER: 60/081,057
PRIOR FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 154
TYPE: PRT
ORGANISM: mouse
US-09-540-245A-10

Query Match 84.0%; Score 42; DB 4; Length 154;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|::|::|::|
Db 11 CMLGYTGDC 20

RESULT 13
US-09-540-153-10
Sequence 10, Application US/09540153
Patent No. 6270995
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions
FILE REFERENCE: B98-031-3
CURRENT APPLICATION NUMBER: US/09/540,153
CURRENT FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/191,647
PRIOR FILING DATE: 1998-11-13
PRIOR APPLICATION NUMBER: 60/081,057
PRIOR FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 10
LENGTH: 154
TYPE: PRT
ORGANISM: mouse
US-09-540-153-10

Query Match 84.0%; Score 42; DB 4; Length 154;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|::|::|::|
Db 11 CMLGYTGDC 20

RESULT 14
5332669-1
Patent No. 5332669
APPLICANT: DEUEL, THOMAS F.
TITLE OF INVENTION: PROSTATE-DERIVED MITOGEN
NUMBER OF SEQUENCES: 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/356,739
FILING DATE: 24-MAR-1989
SEQ ID NO: 1
LENGTH: 48
5332669-1

Query Match 80.0%; Score 40; DB 6; Length 48;
Best Local Similarity 60.0%; Pred. No. 12;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVIGXSGDXC 10
|::|::|::|
Db 33 CVIGYIGERC 42
RESULT 15
US-08-525-864A-2
Sequence 2, Application US/08525864A
Patent No. 5912326
GENERAL INFORMATION:
APPLICANT: Chang, Han
TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses
NUMBER OF SEQUENCES: 18
TITLE OF INVENTION: Related thereto
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,864A
FILING DATE: 8-SEP-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kara, Catherine J.
REGISTRATION NUMBER: 41,106
REFERENCE/DOCKET NUMBER: HUI-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 754 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-864A-2

Query Match 80.0%; Score 40; DB 2; Length 754;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|::|::|::|
Db 280 CPVGYTGDC 289

Search completed: July 2, 2003, 07:25:44
Job time : 12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:28:49 ; Search time 103 Seconds
(without alignments)
11.164 Million cell updates/sec

Title: US-09-673-785D-13
Perfect score: 50
Sequence: 1 CVIGSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 11492915 residues
Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	50	100.0	53	US-10-138-158-19	Sequence 19, Appl
2	50	100.0	53	US-10-150-648B-33	Sequence 33, Appl
3	50	100.0	145	US-10-150-648B-30	Sequence 30, Appl
4	43	86.0	176	US-10-024-599-4	Sequence 4, Appl
5	43	86.0	741	US-09-925-301-930	Sequence 930, Appl
6	42	84.0	572	US-09-900-449A-7	Sequence 7, Appl
7	42	84.0	601	US-09-900-449A-5	Sequence 5, Appl
8	42	84.0	639	US-09-900-449A-4	Sequence 4, Appl
9	40	80.0	112	US-10-211-994-20	Sequence 20, Appl
10	40	80.0	298	US-09-864-675-4	Sequence 4, Appl
11	40	80.0	878	US-10-157-031-171	Sequence 171, Appl
12	39	78.0	46	US-10-136-573A-12	Sequence 12, Appl
13	39	78.0	46	US-09-877-665-12	Sequence 12, Appl
14	39	78.0	46	US-10-215-862-12	Sequence 12, Appl
15	39	78.0	46	US-09-817-647-12	Sequence 12, Appl
16	39	78.0	47	US-10-096-241-17	Sequence 17, Appl
17	39	78.0	48	US-10-201-945-13	Sequence 13, Appl
18	39	78.0	53	US-09-903-327A-9	Sequence 9, Appl
19	39	78.0	53	US-10-150-648B-35	Sequence 35, Appl

Sequence 4, Appl
Sequence 31, Appl
Sequence 3, Appl
Sequence 13, Appl
Sequence 27, Appl
Sequence 29, Appl
Sequence 5, Appl
Sequence 7, Appl
Sequence 25, Appl
Sequence 64, Appl
Sequence 32, Appl
Sequence 692, Appl
Sequence 15, Appl
Sequence 15, Appl
Sequence 5, Appl
Sequence 13, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 42, Appl
Sequence 114, Appl
Sequence 1126, Appl
Sequence 6253, Appl
Sequence 3, Appl
Sequence 2, Appl
Sequence 29, Appl
Sequence 269, Appl

39 78.0 53 9 US-10-211-994-4
39 78.0 53 10 US-09-848-664-31
39 78.0 58 10 US-09-934-706-3
39 78.0 91 9 US-10-022-609-13
39 78.0 111 9 US-10-211-994-27
39 78.0 111 9 US-10-211-994-29
39 78.0 112 9 US-10-211-994-5
39 78.0 112 9 US-10-211-994-7
39 78.0 112 9 US-10-211-994-25
39 78.0 140 10 US-09-280-030-64
39 78.0 146 9 US-10-150-648B-32
39 78.0 255 9 US-09-866-050A-692
39 78.0 275 9 US-10-269-353-15
39 78.0 275 10 US-09-790-264-15
39 78.0 400 10 US-09-934-706-5
39 78.0 493 9 US-09-903-327A-13
39 78.0 832 9 US-09-783-931-6
39 78.0 832 10 US-09-908-322-6
39 78.0 45 10 US-09-995-593A-42
39 76.0 1050 9 US-09-796-753-114
37 74.0 62 10 US-09-867-550-1126
37 74.0 384 9 US-10-106-698-6253
37 74.0 719 9 US-10-133-642-3
37 74.0 2444 10 US-09-944-849-2
36 72.0 16 9 US-10-084-994-29
36 72.0 43 9 US-10-050-704-269

ALIGNMENTS

RESULT 1
US-10-138-158-19
; Sequence 19, Application US/10138158
; Publication No. US20030036509A1
; GENERAL INFORMATION:
; APPLICANT: STEM CELL PHARMACEUTICALS, INC.
; APPLICANT: TWARDZIK, Daniel R.
; APPLICANT: PERNET, Andre
; APPLICANT: FELKER, Thomas S.
; APPLICANT: PASKELL, Stefan
; APPLICANT: RENO, John M.
; TITLE OF INVENTION: TGF-alpha POLYPEPTIDES, FUNCTIONAL FRAGMENTS AND METHODS OF
; FILE REFERENCE: STEM1110-6
; CURRENT APPLICATION NUMBER: US/10/138,158
; CURRENT FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: US 09/641,587
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/559,248
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/459,813
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 09/492,935
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 09/378,567
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-138-158-19

Query Match 100.0%; Score 50; DB 9; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.51;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGSGDXC 10
Db 33 CVIGSGDXC 42

```
RESULT 2
US-10-150-648B-33
; Sequence 33, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725
; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 977 TO 1029
US-10-150-648B-33

Query Match      100.0%; Score 50; DB 9; Length 53;
Best Local Similarity 80.0%; Pred. No. 0.51;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   |||||:|
Db 33 CVIGYSGDRC 42

RESULT 3
US-10-150-648B-30
; Sequence 30, Application US/10150648B
; Publication No. US20030059802A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau-Goesseels, Sylvie
; APPLICANT: John, Sushil J.
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Nucleic acid and protein sequences of bovine epidermal growth
; TITLE OF INVENTION: factor
; FILE REFERENCE: 60-01
; CURRENT APPLICATION NUMBER: US/10/150,648B
; CURRENT FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/292,136
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; AUTHORS: Gray, A., Dull, T.J. and Ullrich, A.
; TITLE: Nucleotide sequence of epidermal growth factor cDNA predicts a
; JOURNAL: Nature
; VOLUME: 303
; PAGES: 722-725
```

```
; DATE: 1983
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030059802A1 J00380
; DATABASE ENTRY DATE: 1993-04-27
; RELEVANT RESIDUES: Relevant residues FROM 919 TO 1063
US-10-150-648B-30

Query Match      100.0%; Score 50; DB 9; Length 145;
Best Local Similarity 80.0%; Pred. No. 1.3;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   |||||:|
Db 91 CVIGYSGDRC 100

RESULT 4
US-10-024-599-4
; Sequence 4, Application US/10024599
; Patent No. US2002016532A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Cimborra, Daniel M.
; APPLICANT: Heinrich, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-278-II
; CURRENT APPLICATION NUMBER: US/10/024,599
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/256,986
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: 1..176
; OTHER INFORMATION: Xaa is Gly or Cys
US-10-024-599-4

Query Match      86.0%; Score 43; DB 9; Length 176;
Best Local Similarity 70.0%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   |||:|:|:|
Db 11 CVNGWSGDRC 20

RESULT 5
US-09-925-301-930
; Sequence 930, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 930
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (282)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-930

Query Match 86.0%; Score 43; DB 10; Length 741;
Best Local Similarity 70.0%; Pred. No. 67;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10
|||:|:|:|
Db 460 CVPYSGDQC 469

RESULT 6

US-09-900-449A-7
; Sequence 7, Application US/09900449A
; Publication No. US20030040616A1
; GENERAL INFORMATION:
; APPLICANT: ZHONG, Jenny et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001271
; CURRENT APPLICATION NUMBER: US/09/900,449A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Strongylocentrotus purpuratus
US-09-900-449A-7

Query Match 84.0%; Score 42; DB 9; Length 572;
Best Local Similarity 60.0%; Pred. No. 76;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10
|||:|:|:|
Db 141 CPLGFGSDNC 150

RESULT 7

US-09-900-449A-5
; Sequence 5, Application US/09900449A
; Publication No. US20030040616A1
; GENERAL INFORMATION:
; APPLICANT: ZHONG, Jenny et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001271
; CURRENT APPLICATION NUMBER: US/09/900,449A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 601
; TYPE: PRT
; ORGANISM: Strongylocentrotus purpuratus
US-09-900-449A-5

Query Match 84.0%; Score 42; DB 9; Length 601;
Best Local Similarity 60.0%; Pred. No. 79;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

Db 132 CPLGFGSDNC 141
|||:|:|:|

RESULT 8

US-09-900-449A-4
; Sequence 4, Application US/09900449A
; Publication No. US20030040616A1
; GENERAL INFORMATION:
; APPLICANT: ZHONG, Jenny et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001271
; CURRENT APPLICATION NUMBER: US/09/900,449A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Strongylocentrotus purpuratus
US-09-900-449A-4

Query Match 84.0%; Score 42; DB 9; Length 639;
Best Local Similarity 60.0%; Pred. No. 84;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10
|||:|:|:|
Db 60 CPLGFGSDNC 69

RESULT 9

US-10-211-994-20
; Sequence 20, Application US/10211994
; Publication No. US20030082201A1
; GENERAL INFORMATION:
; APPLICANT: Rao, M.R.S.
; APPLICANT: Sengupta, Paromita
; APPLICANT: Prasad, Sudhanand
; APPLICANT: Burman, Anand C.
; APPLICANT: Mukherjee, Rama
; APPLICANT: Thomas, Becky
; TITLE OF INVENTION: MULTIVALENT SYNTHETIC VACCINE FOR CANCER
; FILE REFERENCE: U014152-1
; CURRENT APPLICATION NUMBER: US/10/211,994
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: 60/309,975
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide and peptide sequence of multivalent vaccine
US-10-211-994-20

Query Match 80.0%; Score 40; DB 9; Length 112;
Best Local Similarity 60.0%; Pred. No. 35;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10
|||:|:|:|
Db 92 CVIGYIGERC 101

RESULT 10

US-09-864-675-4
; Sequence 4, Application US/09864675
; Patent No. US20020081286A1

GENERAL INFORMATION:
APPLICANT: Marchionni, Mark
TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,
FILE REFERENCE: 04585/049002 POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS
CURRENT APPLICATION NUMBER: US/09/864,675
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/206,495
PRIOR FILING DATE: 2000-05-23
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 298
TYPE: PRT
ORGANISM: Homo sapiens
US-09-864-675-4

Query Match 80.0%; Score 40; DB 10; Length 298;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|:|:|:|:|
Db 280 CPVGYTGDCR 289

RESULT 11

US-10-157-031-171
Sequence 171, Application US/10157031
Publication No. US20030108890A1
GENERAL INFORMATION:

APPLICANT: Baranova, A. V.
APPLICANT: Yankovsky, N. K.
APPLICANT: Kozlov, A. P.
APPLICANT: Lobashev, A. V.
APPLICANT: Krukovskaya, L. L.

TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
FILE REFERENCE: 2760-103
CURRENT APPLICATION NUMBER: US/10/157,031
CURRENT FILING DATE: 2002-05-30
NUMBER OF SEQ ID NOS: 415
SOFTWARE: PatentIn version 3.1
SEQ ID NO 171
LENGTH: 878
TYPE: PRT
ORGANISM: Homo sapiens
US-10-157-031-171

Query Match 80.0%; Score 40; DB 9; Length 878;
Best Local Similarity 60.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|:|:|:|:|
Db 526 CLPGFSGDRC 535

RESULT 12

US-10-136-573A-12
Sequence 12, Application US/10136573A
Patent No. US20020161200A1
GENERAL INFORMATION:

APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie Rose
APPLICANT: Zhang, Dong Xiao

TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
FILE REFERENCE: P1084R1C2
CURRENT APPLICATION NUMBER: US/10/136,573A
CURRENT FILING DATE: 2002-04-29
PRIOR APPLICATION NUMBER: US 09/480,977
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: US 08/899,437

PRIOR FILING DATE: 1997-07-24
PRIOR APPLICATION NUMBER: US 60/052,019
PRIOR FILING DATE: 1997-07-09
NUMBER OF SEQ ID NOS: 23
SEQ ID NO 12
LENGTH: 46
TYPE: PRT
ORGANISM: Homo sapiens
US-10-136-573A-12

Query Match 78.0%; Score 39; DB 9; Length 46;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|:|:|:|:|
Db 32 CVVGYIGERC 41

RESULT 13

US-09-877-665-12
Sequence 12, Application US/09877665
Patent No. US2002016480A1
GENERAL INFORMATION:

APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
Ligands and Uses Therefor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA: US/09/877,665
FILING DATE: 08-Jun-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/109,206
FILING DATE: 30-Jun-1998

ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1-1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear

FEATURE:
NAME/KEY: hEGF.esf
LOCATION: 1-46
IDENTIFICATION METHOD:
OTHER INFORMATION:

SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-877-665-12

Query Match 78.0%; Score 39; DB 9; Length 46;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
|:|:|:|:|

Db 32 CVVGYIGERC 41

RESULT 14

US-10-215-862-12
; Sequence 12, Application US/10215862
; Publication No. US20030036166A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: P1084RID2C1
; CURRENT APPLICATION NUMBER: US/10/215,862
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 09/126,663
; PRIOR FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: US 08/899,437
; PRIOR FILING DATE: 1997-07-24
; PRIOR APPLICATION NUMBER: US 60/052,019
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 12
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-215-862-12

Query Match 78.0%; Score 39; DB 9; Length 46;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

||:|: |::|

Db 32 CVVGYIGERC 41

RESULT 15

US-09-817-647-12
; Sequence 12, Application US/09817647
; Patent No. US20020082229A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/817,647
; FILING DATE: 26-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/107,979
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: heGF.egf
; LOCATION: 1-46
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-817-647-12

Query Match 78.0%; Score 39; DB 10; Length 46;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

||:|: |::|

Db 32 CVVGYIGERC 41

Search completed: July 2, 2003, 07:57:19
Job time : 104 secs

47

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:01 ; Search time 12 seconds
(without alignments)
80.112 Million cell updates/sec

Title: US-09-673-785D-13
Perfect score: 50
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	1217	1 EGSMG	epidermal growth f
2	45	90.0	2139	2 A35672	crumbs protein - f
3	45	90.0	4006	2 T09070	probable tenascin
4	43	86.0	102	2 B55885	chondroitin sulfat
5	43	86.0	372	2 T29359	hypothetical prote
6	43	86.0	862	2 S43922	versican - pig-tai
7	43	86.0	1203	2 A49175	Notch B protein -
8	43	86.0	1643	2 T14274	versican precursor
9	43	86.0	2409	1 A60979	versican precursor
10	43	86.0	2471	2 A49128	cell-fate determin
11	43	86.0	2531	2 T31070	notch homolog - se
12	43	86.0	3381	2 T42389	versican precursor
13	42	84.0	1064	2 A40136	fibropellin Ia - s
14	41	82.0	2352	2 T30201	Notch homolog prot
15	40	80.0	57	2 FC4415	Erbb kinase activa
16	40	80.0	161	2 S03938	neuroendocrine pro
17	40	80.0	601	2 T48539	probable potassium
18	40	80.0	1133	1 EGRT	epidermal growth f
19	40	80.0	1531	2 T42218	slit-1 protein hom
20	40	80.0	2437	2 S42612	transmembrane prot
21	39	78.0	264	2 T16271	hypothetical prote
22	39	78.0	379	2 T16213	APX-1 protein homo
23	39	78.0	422	2 S47581	hypothetical prote
24	39	78.0	426	2 AH1831	hypothetical prote
25	39	78.0	470	2 F85164	hypothetical prote
26	39	78.0	601	2 T34396	hypothetical prote
27	39	78.0	832	2 A31246	neurogenic protein
28	39	78.0	833	2 S19087	gene Delta protein
29	39	78.0	880	2 S00670	neurogenic repetit

30	39	78.0	883	2 S49126	brevican precursor
31	39	78.0	1049	1 S19421	ATP-dependent perm
32	39	78.0	1207	1 EGHU	epidermal growth f
33	39	78.0	1722	2 E89753	protein FilC7.4 (i
34	39	78.0	2397	1 A55535	versican precursor
35	39	78.0	2824	2 T22759	hypothetical prote
36	38	76.0	2476	2 T34022	zonadhesin - pig
37	38	76.0	3566	1 A40701	tenascin-X precurs
38	38	76.0	4135	2 T42629	tenascin-X - bovin
39	38	76.0	4307	2 T20721	hypothetical prote
40	38	76.0	5255	2 T31677	bactracin synthet
41	37	74.0	108	2 A46222	hydrophobin Cey-2
42	37	74.0	294	2 E98236	hypothetical prote
43	37	74.0	294	2 AF3049	conserved hypothet
44	37	74.0	373	2 T22300	hypothetical prote
45	37	74.0	447	2 A39321	mucin - rat (fragm

ALIGNMENTS

RESULT 1

EGSMG

epidermal growth factor precursor - mouse

N:Alternate names: urogastrone precursor

C:Species: Mus musculus (house mouse)

C:Date: 30-Nov-1980 #sequence-revision 11-Aug-1983 #text_change 19-Jan-2001

C:Accession: A94272; A93304; A92118; A01387

R:Scott, J.; Urdea, M.; Quiroga, M.; Sanchez-Pescador, R.; Fong, N.; Selby, M.; Rut

Science 221, 236-240, 1983

A:Title: Structure of a mouse submaxillary messenger RNA encoding epidermal growth

A:Reference number: A94272; MUID:83223630; PMID:6602382

A:Accession: A94272

A:Molecule type: mRNA

A:Residues: 1-1217 <SCO>

A:Cross-references: GB:J00380; NID:g192993; PIDN:AAA37539.1; PID:g309210

R:Gray, A.; Dull, T.; Ullrich, A.

Nature 303, 722-725, 1983

A:Title: Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-mol

A:Reference number: A93304; MUID:83219309; PMID:6304537

A:Accession: A93304

A:Molecule type: mRNA

A:Residues: 1-789,'Y','791-1047','S','1049-1168 <GRA>

A:Cross-references: GB:J00380

A:Note: the sequence shown by these authors differs from residues 1134-1168 due to

ence of Scott et al.)

R:Savage Jr., C.R.; Inagami, T.; Cohen, S.

J. Biol. Chem. 247, 7612-7621, 1972

A:Title: The primary structure of epidermal growth factor.

A:Reference number: A92118; MUID:73048516; PMID:4636327

A:Accession: A92118

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498; PMID:4750422

A:Contents: annotation; disulfide bonds

C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and different

gastrointestinal cell proliferation.

C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in

C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound p

C:Comment: The active growth factor from this submaxillary gland protein stimulates

C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-cr

C:Keywords: duplication; growth factor; tandem repeat; transmembrane protein

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1217/Product: epidermal growth factor proprotein, membrane-bound form #status :

F:29-1038/Domain: extracellular #status predicted <EXT>

F:50-485/Region: EGF precursor long repeat

F:53-92/Domain: LDL receptor YWTD-containing repeat homology <YW01>

F:93-134/Domain: LDL receptor YWTD-containing repeat homology <YW02>

F:135-176/Domain: LDL receptor WYTD-containing repeat homology <YW03>
 F:177-217/Domain: LDL receptor WYTD-containing repeat homology <YW04>
 F:218-262/Domain: LDL receptor WYTD-containing repeat homology <YW05>
 F:263-307/Domain: LDL receptor WYTD-containing repeat homology <YW06>
 F:324-360/Domain: EGF homology #status atypical <EG1>
 F:366-401/Domain: EGF homology <EG2>
 F:407-442/Domain: EGF homology <EG3>
 F:445-482/Domain: EGF homology <EG4>
 F:486-961/Region: EGF precursor long repeat
 F:489-529/Domain: LDL receptor WYTD-containing repeat homology <YW07>
 F:530-572/Domain: LDL receptor WYTD-containing repeat homology <YW08>
 F:573-615/Domain: LDL receptor WYTD-containing repeat homology <YW09>
 F:616-659/Domain: LDL receptor WYTD-containing repeat homology <YW10>
 F:660-700/Domain: LDL receptor WYTD-containing repeat homology <YW11>
 F:701-743/Domain: LDL receptor WYTD-containing repeat homology <YW12>
 F:751-786/Domain: EGF homology <EG5>
 F:842-875/Domain: EGF homology <EG6>
 F:881-917/Domain: EGF homology <EG7>
 F:923-958/Domain: EGF homology <EG8>
 F:977-1029/Product: epidermal growth factor #status experimental <EGF>
 F:982-1018/Domain: EGF homology <EG9>
 F:1039-1063/Domain: transmembrane #status predicted <TMW>
 F:1064-1217/Domain: intracellular #status predicted <INT>
 F:347-360,366-377,373-386,388-401,407-418,414-427,429-442,445-457,453-467,469-482,751-76
 tus predicted
 F:982-996,990-1007,1009-1018/Disulfide bonds: #status experimental

Query Match 100.0%; Score 50; DB 1; Length 1217;
 Best Local Similarity 80.0%; Pred. No. 1.6;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 |||||:
 Db 1009 CVIGSGDRC 1018

RESULT 2

A35672
 crumbs protein - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 21-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 11-Jan-2000
 C:Accession: A35672
 Cell 61, 787-799, 1990
 R:Teppas, U.; Theres, C.; Knust, E.
 A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila
 A:Reference number: A35672; MUID:90263104; PMID:2344615
 A:Accession: A35672
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2139 <TEP>
 A:Cross-references: GB:M33753
 A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue
 C:Genetics:

A:Gene: FlyBase:crb
 A:Cross-references: FlyBase:FBgn0000368
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 C:Keywords: transmembrane protein
 F:352-385/Domain: EGF homology <EGX1>
 F:392-424/Domain: EGF homology <EGF1>
 F:691-722/Domain: EGF homology <EGF2>
 F:767-799/Domain: EGF homology <EGF3>
 F:1878-1914/Domain: EGF homology <EGX2>

Query Match 90.0%; Score 45; DB 2; Length 2139;
 Best Local Similarity 60.0%; Pred. No. 20;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 |||||:
 Db 571 CAVGSGDRC 580

RESULT 3

T09070
 probable tenascin X - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jan-2000
 C:Accession: T09070
 R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.;
 submitted to the EMBL Data Library, October 1997
 A:Description: Sequence of the mouse major histocompatibility locus class III region.
 A:Reference number: Z16543
 A:Accession: T09070
 A:Status: preliminary; translated from GB/EMBL/DBBJ
 A:Molecule type: DNA
 A:Residues: 1-4006 <ROW>
 A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958
 C:Genetics:
 A:Gene: TNX
 A:Map position: 17
 A:introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 144C/1;
 019/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3757/3
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin
 C:Keywords: extracellular matrix
 F:423-448/Domain: EGF homology <EGF>
 F:826-906/Domain: fibronectin type III repeat homology <3FR>
 F:7389-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 90.0%; Score 45; DB 2; Length 4006;
 Best Local Similarity 60.0%; Pred. No. 37;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 |||||:
 Db 532 CAVGSGDDC 541

RESULT 4

B55885
 chondroitin sulfate proteoglycan MV3 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 16-Jul-1999
 C:Accession: B55885
 R:Zako, M.; Shinomura, T.; Ujita, M.; Ito, K.; Kimata, K.
 J. Biol. Chem. 270, 3914-3918, 1995
 A:Title: Expression of PG-M(W3), an alternatively spliced form of PG-M without a cho
 A:Reference number: A35885; MUID:95181355; PMID:7676137
 A:Accession: B55885
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-102 <ZAK>
 A:Cross-references: GB:S75879; GB:D32039
 C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;
 F:24-55/Domain: EGF homology <EG1>
 F:62-93/Domain: EGF homology <EG2>

Query Match 86.0%; Score 43; DB 2; Length 102;
 Best Local Similarity 70.0%; Pred. No. 2.5;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
 |||||:
 Db 46 CVPGYSGDQC 55

RESULT 5

T29359
 hypothetical protein R05G6.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29359
 R:Murray, J.; Le, T.T.
 submitted to the EMBL Data Library, May 1996
 A:Description: The sequence of C. elegans cosmid R05G6.
 A:Reference number: Z20612
 A:Accession: T29359

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-372 <MUR>

A:Cross-references: EMBL:U58746; PIDN:AAB00626.1; GSPDB:GN00022; CESP:R05G6.9

A:Experimental source: strain Bristol N2; clone R05G6

C:Genetics:

A:Gene: CESP:R05G6.9

A:Map position: 4

A:Introns: 80/1; 161/1; 245/1; 286/1

Query Match 86.0%; Score 43; DB 2; Length 372;

Best Local Similarity 50.0%; Pred. No. 8.4;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

|||:|||||

Db 66 CFLGWSGDS 75

RESULT 6

S43922

N:Alternate names: chondroitin sulfate proteoglycan

C:Species: Macaca nemestrina (pig-tailed macaque)

C:Date: 08-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 16-Jul-1999

C:Accession: S43922

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43922

A:Molecule type: mRNA

A:Residues: 1-233;234-525;526-862 <YAO>

A:Cross-references: EMBL:S72413

A:Note: 507-Ser was also found

A:Note: the authors translated the codon GCC for residue 50 as Val, AAG for residue 422

669 as Asn

C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG

C:Keywords: chondroitin sulfate proteoglycan; extracellular matrix

F:1-37/Domain: link protein repeat homology (fragment) <LNK1>

F:58-139/Domain: link protein repeat homology <LNK2>

F:722-753/Domain: EGF homology <EG1>

F:760-791/Domain: EGF homology <EG2>

Query Match 86.0%; Score 43; DB 2; Length 862;

Best Local Similarity 70.0%; Pred. No. 19;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

|||:|||||

Db 744 CVPGYSGDQC 753

RESULT 7

A49175

Motch B protein - mouse (fragment)

N:Alternate names: Notch homolog

C:Species: Mus musculus (house mouse)

C:Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999

C:Accession: A49175; PH1570; S32113

R:Bardele, M.; Lendahl, U.

Exp. Cell Res. 204, 364-372, 1993

A:Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of

A:Reference number: A49175; MUID:93178563; PMID:8440332

A:Accession: A49175

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1203 <LAR>

A:Cross-references: EMBL:X68279; MID:g287989; PIDN:CAA48340.1; PID:g287990

A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBIP:126158)

C:Comment: This protein has many EGF repeats and lin-12/Notch repeats.

C:Comment: This protein is one of the neurogenic proteins controlling the decision betwe

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

F:143-174/Domain: EGF homology <EGX1>

F:482-513/Domain: EGF homology <EGF1>

F:560-591/Domain: EGF homology <EGF>

F:674-705/Domain: EGF homology <EGX2>

F:712-743/Domain: EGF homology <EGF3>

F:836-867/Domain: EGF homology <EGX3>

Query Match 86.0%; Score 43; DB 2; Length 1203;

Best Local Similarity 70.0%; Pred. No. 26;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

|||:|||||

Db 9 CVNGWSGDDC 18

RESULT 8

T14274

versican precursor, splice form V2 - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000

C:Accession: T14274

R:Schmalheid, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.

J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine

A:Reference number: Z17954; MUID:98288320; PMID:9624174

A:Accession: T14274

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1643 <SCH>

A:Cross-references: EMBL:AF060458; MID:g32533303; PID:g32533304; PIDN:AAC24360.1

A:Experimental source: brain

C:Keywords: glycoprotein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-1643/Product: versican, splice form V2 #status predicted <MAT>

F:57,331,352,817,965,1017,1333,1616,1636/Binding site: carbohydrate (Asn) (covalent)

Query Match 86.0%; Score 43; DB 2; Length 1643;

Best Local Similarity 70.0%; Pred. No. 35;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

|||:|||||

Db 1362 CVPGYSGDRC 1371

RESULT 9

A60979

versican precursor - human

N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteogly

N:Contains: glial hyaluronate-binding protein

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179

R:Zimmermann, D.R.; Ruoslahti, E.

EMBO J. 8, 2975-2981, 1989

A:Title: Multiple domains of the large fibroblast proteoglycan, versican.

A:Reference number: S06014; MUID:90059882; PMID:2583089

A:Accession: S06014

A:Molecule type: mRNA

A:Residues: 1-2409 <ZIM>

A:Cross-references: GB:X15998; MID:g37662; PIDN:CAA34128.1; PID:g37663

R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.

Matrix Biol. 14, 213-225, 1994

A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cell

A:Reference number: S43921; MUID:95005762; PMID:7921538

A:Accession: S43921

A:Molecule type: mRNA

A:Residues: 208-440;1094-1385;1910-2246 <YAO>

R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.

Brain Res. Bull. 22, 67-70, 1989

A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilag

A;Reference number: A60979; MUID:89229983; PMID:2469524
A;Accession: A60979
A;Molecule type: protein
A;Residues: 171-210;289-303 <BIG>
R;Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.
J. Biol. Chem. 264, 5981-5987, 1989
A;Title: Isolation and partial characterization of a glial hyaluronate-binding protein.
A;Reference number: A30358; MUID:89174663; PMID:2466833
A;Accession: A30358
A;Molecule type: protein
A;Residues: 24-50;80-87, D', 89-119;128-155;167-218;229-259, 'TR', 261-268;277-283, 'G', 285-
R;Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.
J. Biol. Chem. 262, 13120-13125, 1987
A;Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-like
A;Reference number: A29348; MUID:88007514; PMID:2820964
A;Accession: A29348
A;Molecule type: mRNA
A;Residues: 1725, 'V', 1727-3409 <KRU>
A;Cross-references: GB:J02814
R;Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.
J. Biol. Chem. 267, 23883-23887, 1992
A;Title: Isolation of a large aggregating proteoglycan from human brain.
A;Reference number: A45131; MUID:93054750; PMID:1429726
A;Contents: brain
A;Accession: A45131
A;Molecule type: protein
A;Residues: 21-22, 'X', 24-37 <PE2>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:118884)
R;Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.
Genomics 14, 845-851, 1992
A;Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human chro-
A;Reference number: I54179; MUID:93122792; PMID:1478664
A;Accession: I54179
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 251-347 <RES>
A;Cross-references: GB:S52488; NID:g263313; PIDN:AAB24878.1; PID:g263314
C;Genetics:
A;Gene: GDB:CSPG2
A;Cross-references: GDB:127873; OMIM:118661
A;Map position: 3q12-5q14
C;Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-2409/Product: proteoglycan 24k core protein #status predicted <MAT>
F;167-244/Domain: link protein repeat homology <LNK1>
F;265-346/Domain: link protein repeat homology <LNK2>
F;539-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F;2106-2137/Domain: EGF homology <EG1>
F;2144-2175/Domain: EGF homology <EG2>
F;2182-2302/Domain: C-type lectin homology <LCH>
F;2309-2365/Domain: complement factor H repeat homology <FHD>

Query Match 86.0%; Score 43; DB 1; Length 2409;
Best Local Similarity 70.0%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDQC 10
|||:||||:
Db 2128 CVPGYSGDQC 2137

RESULT 10
A49128
cell-fate determining gene Notch2 protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
A;Accession: A49128
R;Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 116, 931-941, 1992
A;Title: Notch2: a second mammalian Notch gene.
A;Reference number: A49128; MUID:93202015; PMID:1295745
A;Accession: A49128

A;Status: preliminary; not compared with conceptual translation
A;Residues: 1-2471 <WEI>
A;Experimental source: Schwann cell
C;Note: sequence extracted from NCBI backbone (NCBIP:127811)
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;264-295/Domain: EGF homology <EGX1>
F;799-830/Domain: EGF homology <EGF1>
F;877-908/Domain: EGF homology <EGX2>
F;1029-1060/Domain: EGF homology <EGF>
F;1067-1098/Domain: EGF homology <EGX3>
F;1153-1184/Domain: EGF homology <EGF3>
F;1191-1222/Domain: EGF homology <EGX4>
F;1876-1908/Domain: ankyrin repeat homology <AN1>
F;1909-1941/Domain: ankyrin repeat homology <AN2>
F;1943-1975/Domain: ankyrin repeat homology <AN3>
F;1976-2008/Domain: ankyrin repeat homology <AN4>
F;2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 86.0%; Score 43; DB 2; Length 2471;
Best Local Similarity 70.0%; Pred. No. 51;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDQC 10
|||:||||:
Db 326 CVNGWSGDDC 335

RESULT 11
T31070
notch homolog - sea urchin (Lytechinus variegatus)
C;Species: Lytechinus variegatus (Variegated urchin)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C;Accession: T31070
R;Sherwood, D.R.; McClay, D.R.
Development 124, 3363-3374, 1997
A;Title: Identification and localization of a sea urchin Notch homologue: insights in
A;Reference number: Z20966; MUID:97454256; PMID:9310331
A;Accession: T31070
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-2531 <SHE>
A;Cross-references: EMBL:AF000634; NID:g2570350; PID:g2570351; PIDN:AAB2088.1
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 86.0%; Score 43; DB 2; Length 2531;
Best Local Similarity 60.0%; Pred. No. 52;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDQC 10
|||:||||:
Db 615 CPVGTSGDNC 624

RESULT 12
T42389
versican precursor, splice form V0 - bovine
N;Alternate names: chondroitin sulfate proteoglycan
C;Species: Bos primigenius taurus (cattle)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
C;Accession: T42389
R;Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A;Title: Versican V2 is a major extracellular matrix component of the mature bovine
A;Reference number: Z17954; MUID:98288320; PMID:9624174
A;Accession: T42389
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3381 <SCH>
A;Cross-references: EMBL:AF060456; NID:g3253299; PID:g3253300; PIDN:AAC24358.1
C;Superfamily: chicken chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein
C;Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein
F;1-20/Domain: signal sequence #status predicted <SIG>

F:788-819/Domain: EGF homology <EG18>
F:826-857/Domain: EGF homology <EG19>
F:864-895/Domain: EGF homology <EG20>
F:902-933/Domain: EGF homology <EG21>
F:936-1064/Region: avidin-like
F:23-34,28-43,45-54,62-88,180-191,185-200,202-211,218-229,223-238,240-249,256-267,262,57,451-466,468-477,484-495/disulfide bonds: #status predicted
F:489-504,506-515,522-533,537-542,544-553,560-571,565-580,582-591,598-609,603-618,62,63,68,810-819,826-837,831-846,848-857,864-875,869-884,886-895,902-913,907-922,924-933/D

Query Match 84.0%; Score 42; DB 2; Length 1064;
Best Local Similarity 60.0%; Pred. No. 34;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDNC 10
| :|:|:|:|
Db 316 CPLGFGDNC 325

RESULT 14
T30201
Notch homolog protein - sea squirt (Halocynthia roretzi)
C:Species: Halocynthia roretzi
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T30201
R:Hori, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.
Dev. Genes Evol. 207, 371-380, 1997
A:Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the
A:Reference number: 220775
A:Accession: T30201
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2352 <ROR>
A:Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA35571.1
C:Genetics:
A:Gene: Notch

Query Match 82.0%; Score 41; DB 2; Length 2352;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDNC 10
| :|:|:|:|
Db 443 CALGFTGDDC 452

RESULT 15
PC4415
ErbB kinase activator beta, brain and thymus - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Dec-1997 #sequence_revision 10-Dec-1997 #text_change 02-Aug-2002
C:Accession: PC4415
R:Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; M
J. Biochem. 122, 675-680, 1997
A:Title: A novel brain-derived member of the epidermal growth factor family that int
A:Reference number: JC5700; MUID:98006324; PMID:9348101
A:Accession: PC4415
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-57 <RIG>
A:Cross-references: DDBJ:D89997; NID:g2605633; PIDN:BAA23346.1; PID:g2605634
A:Experimental source: PC-12 cell
C:Comment: This protein is a member of the epidermal growth factor family. It is fun
ating the differentiation of MDA-MB-453 cells.
C:Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology
F:1-25/Domain: EGF homology (fragment) <EGF>

Query Match 80.0%; Score 40; DB 2; Length 57;
Best Local Similarity 50.0%; Pred. No. 4.7;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDNC 10
| :|:|:|:|

Wed Jul 2 08:08:18 2003

us-09-673-785d-13.jul1.rpr

Page 6

Db 16 CPVGTGDRC 25

Search completed: July 2, 2003, 07:26:27
Job time : 13 secs

2-100

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 07:25:02 ; Search time 10.3333 Seconds
(without alignments)
40.138 Million cell updates/sec

Title: US-09-673-785D-13
Perfect score: 50
Sequence: 1 CVIGXSGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	100.0	1217	1 EGF_MOUSE	P01132 mus musculus
2	45	90.0	2139	1 CRB_DROME	P10040 drosophila
3	43	86.0	862	1 PGCY_MACNE	P28858 macaca nemo
4	43	86.0	2470	1 NTC2_MOUSE	O35516 mus musculus
5	43	86.0	2471	1 NTC2_HUMAN	O04721 homo sapien
6	43	86.0	2471	1 NTC2_RAT	O9q30 rattus norv
7	43	86.0	3381	1 PGCY_BOVIN	P81282 bos taurus
8	43	86.0	3396	1 PGCY_HUMAN	P13611 homo sapien
9	42	84.0	1064	1 FBPI_STRPU	P10079 strongyloce
10	42	84.0	2920	1 CLR2_MOUSE	Q9r0m0 mus musculus
11	40	80.0	161	1 7B2_XENLA	P18844 xenopus lae
12	40	80.0	756	1 NRG2_MOUSE	P56974 mus musculus
13	40	80.0	1133	1 EGF_RAT	P07522 rattus norv
14	40	80.0	2437	1 NTC1_BRARE	P46530 brachydanio
15	39	78.0	833	1 DL_MOUSE	P10041 drosophila
16	39	78.0	883	1 PGCY_RAT	P55068 rattus norv
17	39	78.0	1049	1 ADP1_YEAST	P25371 saccharomyc
18	39	78.0	1207	1 EGF_HUMAN	P01133 homo sapien
19	39	78.0	1213	1 JAG3_BRARE	Q90y54 brachydanio
20	39	78.0	2319	1 NTC3_RAT	Q9r172 rattus norv
21	39	78.0	2738	1 PGCY_RAT	Q9erb4 rattus norv
22	39	78.0	3358	1 PGCY_MOUSE	Q62059 mus musculus
23	38	76.0	72	1 MT12_MYTED	P80247 mytilus edu
24	38	76.0	636	1 DKS_FALSO	O8xx95 ralstonia s
25	38	76.0	1242	1 JAG1_BRARE	Q9y57 brachydanio
26	38	76.0	2476	1 ZAN_PIG	Q28983 sus scrofa
27	38	76.0	4289	1 TENX_HUMAN	P22105 homo sapien
28	38	76.0	5255	1 BACA_BACLI	O68006 b bacitraci
29	37	74.0	108	1 RODL_NEUCR	Q04571 neurospora
30	37	74.0	738	1 TRFM_MOUSE	P08582 homo sapien
31	37	74.0	738	1 TRFM_MOUSE	Q9r0r1 mus musculus
32	37	74.0	1295	1 GLP1_CAEEL	P21308 caenorhabdi
33	37	74.0	2321	1 NTC3_HUMAN	Q9um47 homo sapien

RESULT 1					ALIGNMENTS				
EGF_MOUSE	ID	EGF_MOUSE	STANDARD;	PRT; 1217 AA.					
AC	P01132;								
DT	21-JUL-1986 (Rel. 01, Created)								
DT	21-JUL-1986 (Rel. 01, Last sequence update)								
DT	15-JUN-2002 (Rel. 41, Last annotation update)								
DE	Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].								
GN	EGF.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_TaxID=10090;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=83223630; PubMed=6602382;								
RA	Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Fong N.M.,								
RA	Selby M., Rutter W.J., Bell G.I.;								
RT	"Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.";								
RL	Science 221:236-240(1983).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=83219309; PubMed=6304537;								
RA	Gray A., Dull T.J., Ullrich A.;								
RT	"Nucleotide sequence of epidermal growth factor cDNA predicts a								
RL	128,000-molecular weight protein precursor.";								
RN	[3]								
RP	SEQUENCE OF 977-1029.								
RX	MEDLINE=73048516; PubMed=4636327;								
RA	Savage C.R. Jr., Inagami T., Cohen S.;								
RT	"The primary structure of epidermal growth factor.";								
RL	J. Biol. Chem. 247:7612-7621(1972).								
RN	[4]								
RP	DISULFIDE BONDS.								
RX	MEDLINE=74025498; PubMed=4750422;								
RA	Savage C.R. Jr., Hash J.H., Cohen S.;								
RT	"Epidermal growth factor. Location of disulfide bonds.";								
RL	J. Biol. Chem. 248:7669-7672(1973).								
RN	[5]								
RP	STRUCTURE BY NMR OF 977-1029.								
RX	MEDLINE=92118798; PubMed=1731873;								
RA	Montellone G.T., Wuehrich K., Burgess A.W., Nice E.C., Wagner G.,								
RT	"Solution structure of murine epidermal growth factor determined by								
RL	NMR spectroscopy and refined by energy minimization with								
RN	[6]								
RP	Biochemistry 31:236-249(1992).								
RX	STRUCTURE BY NMR OF 977-1029.								
RA	MEDLINE=93075811; PubMed=1445923;								
RT	Kohda D., Inagaki F.;								
RL	"Three-dimensional nuclear magnetic resonance structures of mouse								
RT	epidermal growth factor in acidic and physiological pH solutions.";								

P46531 homo sapien
Q21313 caenorhabdi
Q00968 sus scrofa
Q29845 archaeoglob
O14944 homo sapien
P29058 hevea bras
Q25464 mytilus gal
P55028 carassius a
P43256 arabidopsis
P49013 strongyloce
P29057 hevea bras
O64966 gossypium h

Biochemistry 31:11928-11939(1992).
 [7]
 RN STRUCTURE BY NMR OF 980-1024.
 RX MEDLINE-99180407; PubMed-10082370;
 RA Barnham K.J., Torres A.M., Alewood D., Alewood P.F., Domagala T.,
 RA Nice E.C., Norton R.S.;
 RT "Role of the 6-20 disulfide bridge in the structure and activity of
 RT epidermal growth factor";
 RL Protein Sci. 7:1738-1749(1998).
 CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
 CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
 CC FIBROBLASTS IN CELL CULTURE.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 1134
 CC TO 1168 DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J00380; AAA37539.1; -
 CC EMBL; V00741; CAA24115.1; ALT_FRAME.
 CC EMBL; V00741; CAA24116.1; -
 CC PIR; A01387; EGM5MG.
 CC PDB; 1EGF; 31-JAN-94.
 CC PDB; 3EGF; 31-JAN-94.
 CC PDB; 1EPG; 31-JAN-94.
 CC PDB; 1EPH; 31-JAN-94.
 CC PDB; 1EPI; 31-JAN-94.
 CC PDB; 1EPJ; 31-JAN-94.
 CC PDB; 1EPK; 31-JAN-94.
 CC PDB; 1A3P; 29-JUL-98.
 CC MGD; MGI:95290; Egf.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR001336; EGF_1.
 CC InterPro; IPR001861; EGF_Ca.
 CC InterPro; IPR000033; Ldl_receptor_rep.
 CC Pfam; PF00008; EGF; 8.
 CC Pfam; PF00058; ldl_recept_b; 7.
 CC PRINTS; PR00009; EGF_TGF.
 CC SMART; SM00179; EGF_CA; 2.
 CC SMART; SM00001; EGF_like; 7.
 CC SMART; SM00135; LY; 9.
 CC PROSITE; PS00010; ASX_HYDROXYL; 3.
 CC PROSITE; PS00022; EGF_1; 1.
 CC PROSITE; PS01186; EGF_2; 6.
 CC PROSITE; PS01187; EGF_CA; 3.
 KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
 KW Signal; 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 1217
 FT PRO-EPIDERMAL GROWTH FACTOR.
 FT CHAIN 977 1029
 FT EPIDERMAL GROWTH FACTOR.
 FT DOMAIN 29 1038
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1039 1058
 FT POTENTIAL.
 FT DOMAIN 1059 1217
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 327 361
 FT EGF-LIKE 1 (INCOMPLETE).
 FT DOMAIN 362 402
 FT EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 403 443
 FT EGF-LIKE 3.
 FT DOMAIN 441 483
 FT EGF-LIKE 4.
 FT DOMAIN 747 787
 FT EGF-LIKE 5.
 FT DOMAIN 838 876
 FT EGF-LIKE 6.
 FT DOMAIN 877 918
 FT EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 919 959
 FT EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 978 1019
 FT EGF-LIKE 9.
 FT DISULFID 366 377
 FT BY SIMILARITY.
 FT DISULFID 373 386
 FT BY SIMILARITY.
 FT DISULFID 388 401
 FT BY SIMILARITY.
 FT DISULFID 407 418
 FT BY SIMILARITY.

FT DISULFID 414 427 BY SIMILARITY.
 FT DISULFID 429 442 BY SIMILARITY.
 FT DISULFID 445 457 BY SIMILARITY.
 FT DISULFID 453 467 BY SIMILARITY.
 FT DISULFID 469 482 BY SIMILARITY.
 FT DISULFID 751 762 BY SIMILARITY.
 FT DISULFID 758 771 BY SIMILARITY.
 FT DISULFID 773 786 BY SIMILARITY.
 FT DISULFID 842 853 BY SIMILARITY.
 FT DISULFID 847 862 BY SIMILARITY.
 FT DISULFID 864 875 BY SIMILARITY.
 FT DISULFID 881 895 BY SIMILARITY.
 FT DISULFID 888 904 BY SIMILARITY.
 FT DISULFID 906 917 BY SIMILARITY.
 FT DISULFID 923 936 BY SIMILARITY.
 FT DISULFID 930 945 BY SIMILARITY.
 FT DISULFID 947 958 BY SIMILARITY.
 FT DISULFID 982 996 BY SIMILARITY.
 FT DISULFID 990 1007 BY SIMILARITY.
 FT DISULFID 1009 1018 BY SIMILARITY.
 FT DOMAIN 1024 1029 BY SIMILARITY.
 FT CARBOHYD 111 111
 FT CARBOHYD 410 410
 FT CARBOHYD 810 810
 FT CARBOHYD 944 944
 FT CONFLICT 790 790
 FT CONFLICT 1048 1048
 FT STRAND 995 997
 FT STRAND 1006 1008
 FT STRAND 1010 1010
 FT TURN 1011 1012
 FT STRAND 1013 1014
 FT STRAND 1020 1021
 SQ SEQUENCE 1217 AA; 133144 MW; A9C7F3D512F82873 CRC64;
 Query Match 100.0%; Score 50; DB 1; Length 1217;
 Best Local Similarity 80.0%; Pred. No. 0.38;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CVIGXGDXC 10
 Db 1009 CVIGYGDC 1018
 RESULT 2
 CRB_DROME STANDARD; PRT; 2139 AA.
 ID CRB_DROME
 AC P10040;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Crumbs protein precursor (95F).
 GN CRB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-Oregon-R; TISSUE-Embryo;
 RX MEDLINE-90263104; PubMed-2344615;
 RA Tepass U., Theres C., Knust E.;
 RT "Crumbs encodes an EGF-like protein expressed on apical membranes of
 RT Drosophila epithelial cells and required for organization of
 RT epithelia";
 RL Cell 61:787-799(1990).
 RN [2]
 RN SEQUENCE OF 1663-1955 FROM N.A.
 RC TISSUE-Embryo;
 RX MEDLINE-87218537; PubMed-3107986;
 RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,

NOT REQUIRED FOR FULL BIOLOGICAL ACTIVITY
 IN VIVO.

N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 D -> Y (IN REF. 2).
 A -> S (IN REF. 2).

RA Vaessin H., Campos-Ortega J.A.;
 RT "EGF homologous sequences encoded in the genome of Drosophila
 RL melanogaster", and their relation to neurogenic genes.";
 CC EMBO J. 6:761-766(1987).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
 CC POLARITY. IT MAY ACT AS A SIGNAL.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
 CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS
 CC -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M33753; AAA28428.1; ALT_SEQ.
 DR EMBL; X05144; CAA28793.1; -
 DR PIR; B26637; B26637.
 DR PIR; A35672; A35672.
 DR HSP; P00740; 1EDM.
 DR FlyBase; FBgn000368; crb.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 26.
 DR Pfam; PF00054; laminin_G; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR SMART; SM00179; EGF_CA; 16.
 DR SMART; SM00001; EGF-like; 16.
 DR SMART; SM00282; LamG; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 15.
 DR PROSITE; PS00022; EGF_1; 26.
 DR PROSITE; PS01186; EGF_2; 17.
 DR PROSITE; PS01187; EGF_Ca; 12.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 90
 FT CHAIN 91 2139 CROMBS PROTEIN.
 FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2085 2111 POTENTIAL.
 FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 267 303 EGF-LIKE 1.
 FT DOMAIN 306 343 EGF-LIKE 2.
 FT DOMAIN 348 386 EGF-LIKE 3.
 FT DOMAIN 388 425 EGF-LIKE 4.
 FT DOMAIN 427 463 EGF-LIKE 5.
 FT DOMAIN 464 500 EGF-LIKE 6.
 FT DOMAIN 501 532 EGF-LIKE 7.
 FT DOMAIN 545 581 EGF-LIKE 8.
 FT DOMAIN 582 611 EGF-LIKE 9.
 FT DOMAIN 609 646 EGF-LIKE 10.
 FT DOMAIN 648 685 EGF-LIKE 11.
 FT DOMAIN 687 723 EGF-LIKE 12.
 FT DOMAIN 725 761 EGF-LIKE 13.
 FT DOMAIN 763 800 EGF-LIKE 14.
 FT DOMAIN 802 838 EGF-LIKE 15.
 FT DOMAIN 840 902 EGF-LIKE 16.
 FT DOMAIN 904 940 EGF-LIKE 17.
 FT DOMAIN 942 978 EGF-LIKE 18.
 FT DOMAIN 980 1021 EGF-LIKE 19.
 FT DOMAIN 1023 1205 LAMININ G-LIKE 1.
 FT DOMAIN 1207 1243 EGF-LIKE 20.
 FT DOMAIN 1250 1480 LAMININ G-LIKE 2.
 FT DOMAIN 1481 1517 EGF-LIKE 21.

FT DOMAIN 1558 1758 LAMININ G-LIKE 3.
 FT DOMAIN 1759 1795 EGF-LIKE 22.
 FT DOMAIN 1797 1833 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1835 1871 EGF-LIKE 24. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1874 1915 EGF-LIKE 25.
 FT DOMAIN 1915 1951 EGF-LIKE 26.
 FT DOMAIN 1953 1989 EGF-LIKE 27.
 FT DOMAIN 1991 2029 EGF-LIKE 28. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 2030 2070 EGF-LIKE 29. CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 271 282 EGF-LIKE 29.
 FT DISULFID 276 291 BY SIMILARITY.
 FT DISULFID 293 302 BY SIMILARITY.
 FT DISULFID 310 321 BY SIMILARITY.
 FT DISULFID 331 331 BY SIMILARITY.
 FT DISULFID 333 342 BY SIMILARITY.
 FT DISULFID 352 363 BY SIMILARITY.
 FT DISULFID 357 374 BY SIMILARITY.
 FT DISULFID 375 385 BY SIMILARITY.
 FT DISULFID 392 403 BY SIMILARITY.
 FT DISULFID 397 412 BY SIMILARITY.
 FT DISULFID 414 424 BY SIMILARITY.
 FT DISULFID 431 442 BY SIMILARITY.
 FT DISULFID 436 451 BY SIMILARITY.
 FT DISULFID 453 462 BY SIMILARITY.
 FT DISULFID 468 479 BY SIMILARITY.
 FT DISULFID 473 488 BY SIMILARITY.
 FT DISULFID 490 499 BY SIMILARITY.
 FT DISULFID 505 515 BY SIMILARITY.
 FT DISULFID 509 520 BY SIMILARITY.
 FT DISULFID 522 531 BY SIMILARITY.
 FT DISULFID 549 562 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 571 580 BY SIMILARITY.
 FT DISULFID 586 597 BY SIMILARITY.
 FT DISULFID 591 602 BY SIMILARITY.
 FT DISULFID 604 610 BY SIMILARITY.
 FT DISULFID 613 624 BY SIMILARITY.
 FT DISULFID 618 634 BY SIMILARITY.
 FT DISULFID 636 645 BY SIMILARITY.
 FT DISULFID 652 664 BY SIMILARITY.
 FT DISULFID 659 673 BY SIMILARITY.
 FT DISULFID 675 684 BY SIMILARITY.
 FT DISULFID 691 702 BY SIMILARITY.
 FT DISULFID 696 711 BY SIMILARITY.
 FT DISULFID 713 722 BY SIMILARITY.
 FT DISULFID 729 740 BY SIMILARITY.
 FT DISULFID 734 749 BY SIMILARITY.
 FT DISULFID 751 760 BY SIMILARITY.
 FT DISULFID 767 778 BY SIMILARITY.
 FT DISULFID 772 787 BY SIMILARITY.
 FT DISULFID 789 799 BY SIMILARITY.
 FT DISULFID 806 817 BY SIMILARITY.
 FT DISULFID 811 826 BY SIMILARITY.
 FT DISULFID 828 837 BY SIMILARITY.
 FT DISULFID 844 855 BY SIMILARITY.
 FT DISULFID 849 890 BY SIMILARITY.
 FT DISULFID 892 901 BY SIMILARITY.
 FT DISULFID 908 919 BY SIMILARITY.
 FT DISULFID 913 928 BY SIMILARITY.
 FT DISULFID 930 939 BY SIMILARITY.
 FT DISULFID 946 957 BY SIMILARITY.
 FT DISULFID 952 966 BY SIMILARITY.
 FT DISULFID 968 977 BY SIMILARITY.
 FT DISULFID 984 995 BY SIMILARITY.
 FT DISULFID 989 1009 BY SIMILARITY.
 FT DISULFID 1011 1020 BY SIMILARITY.
 FT DISULFID 1211 1222 BY SIMILARITY.
 FT DISULFID 1216 1231 BY SIMILARITY.
 FT DISULFID 1233 1242 BY SIMILARITY.
 FT DISULFID 1485 1496 BY SIMILARITY.
 FT DISULFID 1490 1505 BY SIMILARITY.
 FT DISULFID 1507 1516 BY SIMILARITY.
 FT DISULFID 1763 1774 BY SIMILARITY.

FT DISULFID 1768 1783 BY SIMILARITY.
 FT DISULFID 1785 1794 BY SIMILARITY.
 FT DISULFID 1801 1812 BY SIMILARITY.
 FT DISULFID 1806 1821 BY SIMILARITY.
 FT DISULFID 1823 1832 BY SIMILARITY.
 FT DISULFID 1839 1850 BY SIMILARITY.
 FT DISULFID 1844 1859 BY SIMILARITY.
 FT DISULFID 1861 1870 BY SIMILARITY.
 FT DISULFID 1878 1889 BY SIMILARITY.
 FT DISULFID 1883 1903 BY SIMILARITY.
 FT DISULFID 1905 1914 BY SIMILARITY.
 FT DISULFID 1919 1930 BY SIMILARITY.
 FT DISULFID 1924 1939 BY SIMILARITY.
 FT DISULFID 1941 1950 BY SIMILARITY.
 FT DISULFID 1957 1968 BY SIMILARITY.
 FT DISULFID 1962 1977 BY SIMILARITY.
 FT DISULFID 1979 1988 BY SIMILARITY.
 FT DISULFID 1995 2008 BY SIMILARITY.
 FT DISULFID 2002 2017 BY SIMILARITY.
 FT DISULFID 2019 2028 BY SIMILARITY.
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 Query Match 90.0%; Score 45; DB 1; Length 2139;
 Best Local Similarity 60.0%; Pred. No. 5.2;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CVIGXSGDNC 10
 Db 571 CAVGYSGDRC 580
 RESULT 3
 PGCV_MACNE
 ID PGCV_MACNE STANDARD; PRT; 862 AA.
 AC Q28858; Q28859; Q28860;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Versican core protein (large fibroblast proteoglycan) (Chondroitin
 sulfate proteoglycan core protein 2) (Fragments).
 GN CSF62.
 OS Macaca nemestrina (Pig-tailed macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9345;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Aortic smooth muscle;
 RX MEDLINE=95005762; PubMed=7921538;
 RA Yao L.Y., Moody C., Schoenherr E., Wight T.N., Sandell L.J.;
 RT "Identification of the proteoglycan versican in aorta and smooth
 muscle cells by DNA sequence analysis, in situ hybridization and
 immunohistochemistry.";
 RL Matrix Biol. 14:213-225(1994).
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 connecting cells with the extracellular matrix. May take part in
 the regulation of cell motility, growth and differentiation. Binds
 hyaluronan.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
 similarity).
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
 (By similarity).
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.

CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S72412; AAA65593.2; -
 DR EMBL; S72413; AAA65594.2; -
 DR EMBL; S72414; AAA65595.2; -
 DR HSP; P01132; EGF.
 DR InterPro; IPR00152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR PRINTS; PR00010; EGFBLD.
 DR PRODOM; PD000918; Link; 2.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00001; EGF-like; 1.
 DR SMART; SM00445; LINK; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS01241; LINK; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; PARTIAL.
 DR PROSITE; PS00041; C-TYPE LECTIN_2; 1
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Repeat;
 KW EGF-like domain; Calcium.
 FT NON_TER 1
 FT DOMAIN <1 37 LINK 1.
 FT DOMAIN 58 139 LINK 2.
 FT DOMAIN 141 >233 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
 DOMAIN).
 FT NON_CONS 233 234 GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN.
 FT DOMAIN <234 >525 SIMILAR TO CHONDROITIN SULFATE ATTACHMENT
 SITE IN COLLAGEN TYPE IX (BY SIMILARITY).
 FT NON_CONS 525 526 EGF-LIKE 1.
 FT DOMAIN 718 754 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 756 792 C-TYPE LECTIN.
 FT DOMAIN 805 >862 BY SIMILARITY.
 FT DISULFID 63 138 BY SIMILARITY.
 FT DISULFID 87 108 BY SIMILARITY.
 FT DISULFID 722 733 BY SIMILARITY.
 FT DISULFID 727 742 BY SIMILARITY.
 FT DISULFID 744 753 BY SIMILARITY.
 FT DISULFID 760 771 BY SIMILARITY.
 FT DISULFID 765 780 BY SIMILARITY.
 FT DISULFID 782 791 BY SIMILARITY.
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 563 563 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 862 862
 SQ SEQUENCE 862 AA; 95583 MW; A5D5F6153A74BB39 CRC64;
 Query Match 86.0%; Score 43; DB 1; Length 862;
 Best Local Similarity 70.0%; Pred. No. 4.9;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CVIGXSGDNC 10


```

Db          744 CVPYSGDOC 753
RESULT 4
NTC2_MOUSE
ID NTC2_MOUSE STANDARD; PRT: 2470 AA.
AC O35516; Q60941; Q06008;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Motch.
DE B).
DE NOTCH2.
GN Mus musculus (Mouse).
OS
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RA Hamada Y., Higuchi M., Tsujimoto Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
RT single copy of mouse Notch2 gene.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [3]
RN SEQUENCE OF 1765-2153 FROM N.A.
RX MEDLINE=97075110; PubMed=8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
RN [4]
RN FUNCTION.
RX MEDLINE=99396706; PubMed=10393120;
RA Hamada Y., Kadohawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA Tsujimoto Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT embryonic lethality.";
RL Development 126:3415-3424(1999).
RN [5]
RN DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX MEDLINE=95333893; PubMed=7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
RT mouse brain.";
RL Brain Res. Mol. Brain Res. 29:263-272(1995).
RN [6]
RN POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [7]
RN POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with

```

```

CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC neuroepithelia, somites, optic vesicles and branchial arches, but
CC not heart.
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC the postnatal ependymal cells, and the choroid plexus throughout
CC embryonic and postnatal development.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D32210; BAA22094.1; -
CC EMBL; X68279; CAA48340.1; -
CC EMBL; U31881; AAC52924.1; -
CC HSP; P16109; IFSB.
CC MGD; MGI:97364; Notch2.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR001438; EGF_II.
CC InterPro; IPR000800; Notch.
CC Pfam; PF00008; EGF; 35.
CC Pfam; PF00023; ank; 6.
CC Pfam; PF00066; notch; 2.
CC PRINTS; PR00010; EGFBL00D.
CC PRINTS; PR01452; NOTCH.
CC SMART; SM00248; ANK; 4.
CC SMART; SM00179; EGF_CA; 22.
CC SMART; SM00001; EGF_like; 12.
CC SMART; SM00004; NL; 3.
CC PROSITE; PSS0088; ANK_REPEAT; 4.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS00010; ASX_HYDROXYL; 22.
CC PROSITE; PS00022; EGF_1; 33.
CC PROSITE; PS01186; EGF_2; 27.
CC PROSITE; PS01187; EGF_CA; 22.
CC Receptor; Transcription regulation; Activator; Differentiation;
CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
CC Transmembrane; Glycoprotein; Signal; Phosphorylation;

```


notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (by similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.

-!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.

-!- SIMILARITY: CONTAINS 6 ANK REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF308601; AAA36377.2; -.

EMBL; AF315356; AAG37073.1; -.

EMBL; U77493; AAB19224.1; -.

HSSP; P00740; 1EDM.

Genew; HGNC:7882; NOTCH2.

MIM; 600275; -.

InterPro; IPR002110; ANK.

InterPro; IPR000152; Asx_hydroxyl.

InterPro; IPR000561; EGF-like.

InterPro; IPR000742; EGF-2.

InterPro; IPR001881; EGF-Ca.

InterPro; IPR001438; EGF-II.

InterPro; IPR002049; Laminin_EGF.

InterPro; IPR000800; Notch.

Pfam; PF00008; EGF; 35.

Pfam; PF00023; ank; 6.

Pfam; PF00066; notch; 2.

PRINTS; PR00010; EGF_BLOOD.

PRINTS; PR00011; EGF_LAMININ.

SMART; SM00248; ANK; 4.

SMART; SM00179; EGF_Ca; 22.

SMART; SM00001; EGF-like; 12.

SMART; SM00004; ND; 2.

PROSITE; PS50088; ANK_REPEAT; 4.

PROSITE; PS50297; ANK_REPEAT_REGION; 1.

PROSITE; PS00010; ASX_HYDROXYL; 22.

PROSITE; PS00022; EGF_1; 34.

PROSITE; PS01186; EGF_2; 29.

PROSITE; PS01187; EGF_Ca; 22.

Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.

SIGNAL 1 25

CHAIN 26 2471

CHAIN 1666 2471

CHAIN 1697 2471

DOMAIN 26 1677

TRANSMEM 1678 1698

DOMAIN 1699 2471

DOMAIN 26 63

DOMAIN 64 102

DOMAIN 105 143

DOMAIN 144 180

DOMAIN 182 219

DOMAIN 221 238

DOMAIN 260 296

DOMAIN 298 336

DOMAIN 338 374

DOMAIN 375 413

DOMAIN 415 454

DOMAIN 456 492

DOMAIN 494 530

DOMAIN 532 568

notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (by similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.

-!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.

-!- SIMILARITY: CONTAINS 6 ANK REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF308601; AAA36377.2; -.

EMBL; AF315356; AAG37073.1; -.

EMBL; U77493; AAB19224.1; -.

HSSP; P00740; 1EDM.

Genew; HGNC:7882; NOTCH2.

MIM; 600275; -.

InterPro; IPR002110; ANK.

InterPro; IPR000152; Asx_hydroxyl.

InterPro; IPR000561; EGF-like.

InterPro; IPR000742; EGF-2.

InterPro; IPR001881; EGF-Ca.

InterPro; IPR001438; EGF-II.

InterPro; IPR002049; Laminin_EGF.

InterPro; IPR000800; Notch.

Pfam; PF00008; EGF; 35.

Pfam; PF00023; ank; 6.

Pfam; PF00066; notch; 2.

PRINTS; PR00010; EGF_BLOOD.

PRINTS; PR00011; EGF_LAMININ.

SMART; SM00248; ANK; 4.

SMART; SM00179; EGF_Ca; 22.

SMART; SM00001; EGF-like; 12.

SMART; SM00004; ND; 2.

PROSITE; PS50088; ANK_REPEAT; 4.

PROSITE; PS50297; ANK_REPEAT_REGION; 1.

PROSITE; PS00010; ASX_HYDROXYL; 22.

PROSITE; PS00022; EGF_1; 34.

PROSITE; PS01186; EGF_2; 29.

PROSITE; PS01187; EGF_Ca; 22.

Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal; Phosphorylation.

SIGNAL 1 25

CHAIN 26 2471

CHAIN 1666 2471

CHAIN 1697 2471

DOMAIN 26 1677

TRANSMEM 1678 1698

DOMAIN 1699 2471

DOMAIN 26 63

DOMAIN 64 102

DOMAIN 105 143

DOMAIN 144 180

DOMAIN 182 219

DOMAIN 221 238

DOMAIN 260 296

DOMAIN 298 336

DOMAIN 338 374

DOMAIN 375 413

DOMAIN 415 454

DOMAIN 456 492

DOMAIN 494 530

DOMAIN 532 568

FT	DOMAIN	570	605	EGF-LIKE 15,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	607	643	EGF-LIKE 16,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	645	680	EGF-LIKE 17,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	682	718	EGF-LIKE 18,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	720	755	EGF-LIKE 19.	
FT	DOMAIN	757	793	EGF-LIKE 20,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	795	831	EGF-LIKE 21,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	833	871	EGF-LIKE 22.	
FT	DOMAIN	873	909	EGF-LIKE 23,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	911	947	EGF-LIKE 24,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	949	985	EGF-LIKE 25,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	987	1023	EGF-LIKE 26,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1025	1061	EGF-LIKE 27,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1063	1099	EGF-LIKE 28.	
FT	DOMAIN	1101	1147	EGF-LIKE 29.	
FT	DOMAIN	1149	1185	EGF-LIKE 30,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1187	1223	EGF-LIKE 31,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1225	1262	EGF-LIKE 32,	
FT	DOMAIN	1264	1302	EGF-LIKE 33.	
FT	DOMAIN	1304	1343	EGF-LIKE 34.	
FT	DOMAIN	1374	1412	EGF-LIKE 35.	
FT	REPEAT	1420	1456	LIN/NOTCH 1.	
FT	REPEAT	1503	1535	LIN/NOTCH 2.	
FT	REPEAT	1827	1871	ANK 1.	
FT	REPEAT	1876	1905	ANK 2.	
FT	REPEAT	1909	1939	ANK 3.	
FT	REPEAT	1943	1972	ANK 4.	
FT	REPEAT	1976	2005	ANK 5.	
FT	REPEAT	2009	2038	ANK 6.	
FT	DOMAIN	1645	1648	POLY-ALA.	
FT	DOMAIN	1994	1997	POLY-LEU.	
FT	DOMAIN	2426	2429	POLY-SER.	
FT	DISULFID	28	41	BY SIMILARITY.	
FT	DISULFID	35	51	BY SIMILARITY.	
FT	DISULFID	53	62	BY SIMILARITY.	
FT	DISULFID	68	79	BY SIMILARITY.	
FT	DISULFID	73	90	BY SIMILARITY.	
FT	DISULFID	92	101	BY SIMILARITY.	
FT	DISULFID	109	121	BY SIMILARITY.	
FT	DISULFID	115	131	BY SIMILARITY.	
FT	DISULFID	133	142	BY SIMILARITY.	
FT	DISULFID	148	159	BY SIMILARITY.	
FT	DISULFID	153	168	BY SIMILARITY.	
FT	DISULFID	170	179	BY SIMILARITY.	
FT	DISULFID	186	198	BY SIMILARITY.	
FT	DISULFID	192	207	BY SIMILARITY.	
FT	DISULFID	209	218	BY SIMILARITY.	
FT	DISULFID	225	236	BY SIMILARITY.	
FT	DISULFID	230	246	BY SIMILARITY.	
FT	DISULFID	248	257	BY SIMILARITY.	
FT	DISULFID	264	275	BY SIMILARITY.	
FT	DISULFID	269	284	BY SIMILARITY.	
FT	DISULFID	286	295	BY SIMILARITY.	

Query Match 86.0%; Score 43; DB 1; Length 2471;

Best Local Similarity 70.0%; Pred. No. 14;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10

Db 326 CVNGWSGDDC 335

|||:||||

RESULT 6

NTC2_RAT

ID NTC2_RAT STANDARD; PRT; 2471 AA.

AC Q9W30;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 13-JUN-2002 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).

GN NOTCH2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE-93202015; PubMed-1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RL "Notch2: a second mammalian Notch gene.";
 RT Development 116:931-941(1992).
 RN [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE-21331789; PubMed-11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 functional roles for the Notch-DSL signaling system during brain
 development.";
 RT J. Comp. Neurol. 436:167-181(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May play an essential role in postimplantation
 CC development, probably in some aspect of cell specification and/or
 CC differentiation (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
 CC plexus in the brain. Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within numerous
 CC cells throughout the CNS. It is more highly localized to
 CC ventricular germinal zones. Also found in the heart, liver and
 CC kidney.
 CC -!- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXI). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M93661; AAK13558.1; --
 DR HSSP; P00743; ICCF.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR002049; Laminin_EGF.

DR InterPro: IPR000800; Notch.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBLLOOD.
 DR PRINTS; PR00011; EGFAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 4.
 DR SMART; SM00179; EGF_CA; 22.
 DR SMART; SM00001; EGF_Like; 10.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS00088; ANK_REPEAT; 4.
 DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 2471
 FT CHAIN 1666 2471
 FT CHAIN 1697 2471
 FT DOMAIN 26 1677
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2471
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 180
 FT DOMAIN 182 219
 FT DOMAIN 221 258
 FT DOMAIN 260 296
 FT DOMAIN 298 336
 FT DOMAIN 338 374
 FT DOMAIN 375 413
 FT DOMAIN 415 454
 FT DOMAIN 456 492
 FT DOMAIN 494 530
 FT DOMAIN 532 568
 FT DOMAIN 570 605
 FT DOMAIN 607 643
 FT DOMAIN 645 680
 FT DOMAIN 682 718
 FT DOMAIN 720 755
 FT DOMAIN 757 793
 FT DOMAIN 795 831
 FT DOMAIN 833 871
 FT DOMAIN 873 909
 FT DOMAIN 911 947
 FT DOMAIN 949 985
 FT DOMAIN 987 1023
 FT DOMAIN 1025 1061
 FT DOMAIN 1063 1099
 FT DOMAIN 1101 1147
 FT DOMAIN 1149 1185
 FT DOMAIN 1187 1223
 FT DOMAIN 1225 1262
 FT DOMAIN 1264 1302
 FT DOMAIN 1304 1343
 FT DOMAIN 1374 1412
 FT DOMAIN 1412 1648
 FT DOMAIN 1645 1664
 FT DOMAIN 1694 1997
 FT DOMAIN 2426 2429
 FT DOMAIN 2446 2451
 FT REPEAT 1420 1456
 FT REPEAT 1503 1535
 FT REPEAT 1827 1871
 FT REPEAT 1876 1905
 FT REPEAT 1909 1939

```

FT REPEAT 1943 1972 ANK 4.
FT REPEAT 1976 2005 ANK 5.
FT REPEAT 2009 2038 ANK 6.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 90 BY SIMILARITY.
FT DISULFID 92 101 BY SIMILARITY.
FT DISULFID 109 121 BY SIMILARITY.
FT DISULFID 115 131 BY SIMILARITY.
FT DISULFID 133 142 BY SIMILARITY.
FT DISULFID 148 159 BY SIMILARITY.
FT DISULFID 153 168 BY SIMILARITY.
FT DISULFID 170 179 BY SIMILARITY.
FT DISULFID 186 198 BY SIMILARITY.
FT DISULFID 192 207 BY SIMILARITY.
FT DISULFID 209 218 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 230 246 BY SIMILARITY.
FT DISULFID 248 257 BY SIMILARITY.
FT DISULFID 264 275 BY SIMILARITY.
FT DISULFID 269 284 BY SIMILARITY.
FT DISULFID 286 295 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 309 324 BY SIMILARITY.
FT DISULFID 326 335 BY SIMILARITY.
FT DISULFID 342 353 BY SIMILARITY.
FT DISULFID 347 362 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 384 401 BY SIMILARITY.
FT DISULFID 403 412 BY SIMILARITY.
FT DISULFID 419 433 BY SIMILARITY.
FT DISULFID 427 442 BY SIMILARITY.
FT DISULFID 444 453 BY SIMILARITY.
FT DISULFID 460 471 BY SIMILARITY.
FT DISULFID 465 480 BY SIMILARITY.
FT DISULFID 482 491 BY SIMILARITY.
FT DISULFID 498 509 BY SIMILARITY.
FT DISULFID 503 518 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 536 547 BY SIMILARITY.
FT DISULFID 541 556 BY SIMILARITY.
FT DISULFID 558 567 BY SIMILARITY.
FT DISULFID 574 584 BY SIMILARITY.

Query Match 86.0%; Score 43; DB 1; Length 2471;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 326 CVNGWSGDDC 335

RESULT 7
PCGV_BOVIN STANDARD; PRT: 3381 AA.
ID P81282; 077609; 077610; 077611; 077612;
AC 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia1
DE hyaluronate-binding protein) (GHP).
GN CP5G2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORMS V0; V1; V2 AND V3).
RC TISSUE=Forebrain;
RX MEDLINE=98288320; PubMed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RC TISSUE=Spinal cord;
RX MEDLINE=92062692; PubMed=1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid.";
RL Biochim. Biophys. Acta 1075:248-258(1991).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2
CC and V3; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF060456; AAC24358.1; -
DR EMBL; AF060457; AAC24359.1; -
DR EMBL; AF060458; AAC24360.1; -
DR EMBL; AF060459; AAC24361.1; -
DR HSSP; P01132; IEPG.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003600; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; XLINK; 2.
DR PRINTS; PR01265; LINKMODULE.
DR PRODOM; PD000918; Link; 1.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG_like; 1.

```

DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN; 1.
 DR PROSITE; PS00411; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_2; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 21 3381
 FT DOMAIN 37 138
 FT DOMAIN 168 245
 FT DOMAIN 266 347
 FT DOMAIN 349 1336
 FT
 FT DOMAIN 1337 3074
 FT DOMAIN 3074 3110
 FT DOMAIN 3112 3148
 FT DOMAIN 3161 3275
 FT DOMAIN 3280 3338
 FT
 FT DISULFID 44 131
 FT DISULFID 173 244
 FT DISULFID 197 218
 FT DISULFID 271 346
 FT DISULFID 295 316
 FT DISULFID 3078 3089
 FT DISULFID 3083 3098
 FT DISULFID 3100 3109
 FT DISULFID 3116 3127
 FT DISULFID 3121 3136
 FT DISULFID 3138 3147
 FT DISULFID 3154 3165
 FT DISULFID 3250 3266
 FT DISULFID 3281 3324
 FT DISULFID 3310 3337
 FT CARBOHYD 57 57
 FT CARBOHYD 331 331
 FT CARBOHYD 352 352
 FT CARBOHYD 817 817
 FT CARBOHYD 965 965
 FT CARBOHYD 1017 1017
 FT CARBOHYD 1333 1333
 FT CARBOHYD 1393 1393
 FT CARBOHYD 1437 1437
 FT CARBOHYD 1463 1463
 FT CARBOHYD 1653 1653
 FT CARBOHYD 1974 1974
 FT CARBOHYD 2045 2045
 FT CARBOHYD 2074 2074
 FT CARBOHYD 2103 2103
 FT CARBOHYD 2263 2263
 FT CARBOHYD 2290 2290
 FT CARBOHYD 2356 2356
 FT CARBOHYD 2623 2623
 FT CARBOHYD 2641 2641
 FT CARBOHYD 2919 2919
 FT CARBOHYD 3052 3052
 FT CARBOHYD 3354 3354
 FT CARBOHYD 3364 3364
 FT VARSPLIC 349 349
 FT VARSPLIC 350 1336
 FT VARSPLIC 1337 3074
 FT VARSPLIC 350 3074
 FT CONFLICT 25 25
 FT CONFLICT 51 51
 FT CONFLICT 89 89
 FT CONFLICT 96 96
 FT CONFLICT 346 346
 FT CONFLICT 3381 AA; 369984 MW; F09716FA7778D459 CRC64;
 SQ SEQUENCE

Query Match 86.0%; Score 43; DB 1; Length 3381;
 Best Local Similarity 70.0%; Pred. No. 19;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CVIGSGDXC 10
 DB 3100 CVPYSGDRC 3109
 |||:|:|

RESULT 8

PGCV_HUMAN
 ID PGCV_HUMAN STANDARD; PRT; 3396 AA.
 AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)
 DE hyaluronate-binding protein) (GHP).
 GN CSPG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM V0).
 RX MEDLINE=95105188; PubMed=7528742;
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
 RT "Characterization of the complete genomic structure of the human
 versican gene and functional analysis of its promoter.";
 RL J. Biol. Chem. 269:32999-33008(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V1).
 RC TISSUE=Placenta;
 RX MEDLINE=90059882; PubMed=2583089;
 RA Zimmermann D.R., Ruoslahti E.;
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";
 RL EMBO J. 8:2975-2981(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM V2).
 RC TISSUE=Glial tumor;
 RX MEDLINE=95105187; PubMed=7806529;
 RA Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "A novel glycosaminoglycan attachment domain identified in two
 alternative splice variants of human versican.";
 RL J. Biol. Chem. 269:32992-32998(1994).
 RN [4]
 RP SEQUENCE OF 2711-3396 FROM N.A.
 RC TISSUE=Lung fibroblast;
 RX MEDLINE=88007514; PubMed=2820964;
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 lectin-like and growth factor-like sequences.";
 RL J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE=93122792; PubMed=1478664;
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McPherson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 human chromosome 5 (5q12-5q14).";
 RL Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RC TISSUE=Brain;
 RX MEDLINE=95181355; PubMed=7876137;
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 without a chondroitin sulfate attachment in region in mouse and human
 tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [7]

RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RC TISSUE-Aortic smooth muscle;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
R Wright T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [8]
RP PARTIAL SEQUENCE.
RC TISSUE-Brain;
RX MEDLINE=89174663; PubMed=2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial
hyaluronate-binding protein.";
RL J. Biol. Chem. 264:5981-5987(1989).
RN [9]
RP TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE=96213482; PubMed=8627343;
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT "Differential expression of versican isoforms in brain tumors.";
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
CC -|- FUNCTION: May play a role in intercellular signaling and in
connecting cells with the extracellular matrix. May take part in
the regulation of cell motility, growth and differentiation. Binds
hyaluronic acid.
CC -|- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -|- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
V2, V3 and Vint; are produced by alternative splicing.
CC -|- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
in normal brain, gliomas, medulloblastomas, schwannomas,
neurofibromas, and meningiomas; V2 is restricted to normal brain
and gliomas; V3 is found in all these tissues except
medulloblastomas.
CC -|- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U16306; AAA65018.1; -;
DR EMBL; X15998; CAA34128.1; -;
DR EMBL; S52488; AAB24878.1; -;
DR EMBL; U26555; AAA67565.1; -;
DR EMBL; D32039; BAA06801.1; -;
DR EMBL; J02814; AAA36437.1; -;
DR EMBL; AF084545; AAD48545.1; -;
DR PIR; S06014; S06014.
DR PIR; A29348; A29348.
DR PIR; A30358; A30358.
DR HSP; P01132; IEGF.
DR HSP; HGNC:2464; CSPG2.
DR MIM; 118661; -;
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.

DR Pfam; PF00047; ig; 1.
DR Pfam; PF00059; lectin_C; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBL00D.
DR PRODOM; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3396 VERSICAN CORE PROTEIN.
FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 167 244 LINK 1.
FT DOMAIN 265 346 LINK 2.
FT DOMAIN 348 1335 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
DOMAIN).
FT DOMAIN 1336 3089 GAG-BETA.
FT DOMAIN 3089 3125 EGF-LIKE 1.
FT DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3176 3290 C-TYPE LECTIN.
FT DOMAIN 3295 3353 SUSHI.
FT DISULFID 44 130 BY SIMILARITY.
FT DISULFID 172 243 BY SIMILARITY.
FT DISULFID 196 217 BY SIMILARITY.
FT DISULFID 270 345 BY SIMILARITY.
FT DISULFID 294 315 BY SIMILARITY.
FT DISULFID 3093 3104 BY SIMILARITY.
FT DISULFID 3098 3113 BY SIMILARITY.
FT DISULFID 3115 3124 BY SIMILARITY.
FT DISULFID 3131 3142 BY SIMILARITY.
FT DISULFID 3136 3151 BY SIMILARITY.
FT DISULFID 3153 3162 BY SIMILARITY.
FT DISULFID 3169 3180 BY SIMILARITY.
FT DISULFID 3197 3289 BY SIMILARITY.
FT DISULFID 3265 3281 BY SIMILARITY.
FT DISULFID 3296 3339 BY SIMILARITY.
FT DISULFID 3325 3352 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 809 809 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1442 1442 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1663 1663 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2179 2179 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2280 2280 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2360 2360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 86.0%;

Best Local Similarity 70.0%;

Pred. No. 19;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10

|||:|:|:|

Db 3115 CVPGYSGDQC 3124

RESULT 9
FBP1_STRPU STANDARD; PRT; 1064 AA.

AC P10079;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibropellin I precursor (Epidermal growth factor-related protein 1)
DE (UEGF-1).
GN EGF1.

OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90112459; PubMed=2514273;
RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;
RT "Structural analysis of the uegf gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats.";
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE=87319677; PubMed=3498216;
RA Hursh D.A., Andrews M.E., Raff R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor.";
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN
RX MEDLINE=89196806; PubMed=2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin.";
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91285254; PubMed=2060714;
RA Bisgrove B.W., Andrews M.E., Raff R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo.";
RL Dev. Biol. 146:89-99(1991).
CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF
CC REPEATS.
CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.

or send an email to license@isb-sib.ch).

EMBL; L08692; AAA62164.1; -
EMBL; L08692; AAA62163.1; -
EMBL; X17530; CAA35571.1; -
EMBL; M17421; AAA30050.1; -
EMBL; X17533; CAA35573.1; -
PIR; A29316; A29316.
HSSP; P01132; LEGF.
InterPro; IPR000152; Asx_hydroxyl.
InterPro; IPR000088; Avidin.
InterPro; IPR000859; CUB_domain.
InterPro; IPR000561; EGF-like.
InterPro; IPR000742; EGF_2.
InterPro; IPR001881; EGF_Ca.
InterPro; IPR001438; EGF_II.
InterPro; IPR002049; Laminin_EGF.
Pfam; PF00008; EGF; 21.
Pfam; PF00431; CUB; 1.
Pfam; PF01382; Avidin; 1.
PRINTS; PR00010; EGFBL00D.
PRINTS; PR00011; EGF_LAMININ.
SMART; SM00042; CUB; 1.
SMART; SM00179; EGF_CA; 18.
SMART; SM00001; EGF_like; 3.
PROSITE; PS00010; ASX_HYDROXYL; 19.
PROSITE; PS00022; EGF_1; 19.
PROSITE; PS00577; AVIDIN; 1.
PROSITE; PS01180; CUB; 1.
PROSITE; PS01186; EGF_2; 19.
PROSITE; PS01187; EGF_CA; 18.
KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
KW Glycoprotein; Calcium-binding.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1064 FIBROPELLIN I.
FT DOMAIN 20 55 EGF-LIKE 1.
FT DOMAIN 62 175 CUB.
FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 860 896 EGF-LIKE 20.
FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 936 1064 AVIDIN-LIKE.
FT DISULFID 23 34 BY SIMILARITY.
FT DISULFID 28 43 BY SIMILARITY.
FT DISULFID 45 54 BY SIMILARITY.
FT DISULFID 180 191 BY SIMILARITY.
FT DISULFID 185 200 BY SIMILARITY.
FT DISULFID 202 211 BY SIMILARITY.
FT DISULFID 218 229 BY SIMILARITY.
FT DISULFID 223 238 BY SIMILARITY.
FT DISULFID 240 249 BY SIMILARITY.
FT DISULFID 256 267 BY SIMILARITY.
FT DISULFID 261 276 BY SIMILARITY.
FT DISULFID 278 287 BY SIMILARITY.
FT DISULFID 294 305 BY SIMILARITY.
FT DISULFID 299 314 BY SIMILARITY.
FT DISULFID 316 325 BY SIMILARITY.


```
FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 337 352 BY SIMILARITY.
FT DISULFID 354 363 BY SIMILARITY.
FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 603 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 750 761 BY SIMILARITY.
FT DISULFID 755 770 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 793 808 BY SIMILARITY.
FT DISULFID 810 819 BY SIMILARITY.
FT DISULFID 826 837 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 848 857 BY SIMILARITY.
FT DISULFID 864 875 BY SIMILARITY.
FT DISULFID 869 884 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 902 913 BY SIMILARITY.
FT DISULFID 907 922 BY SIMILARITY.
FT DISULFID 924 933 BY SIMILARITY.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 477 780 MISSING (IN ISOFORM IB).
FT CONFLICT 279 279 L -> S (IN REF. 2).
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6DD09 CRC64;

Query Match 84.0%; Score 42; DB 1; Length 1064;
Best Local Similarity 60.0%; Pred. No. 9.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
DB 316 CPLGFGDNC 325

RESULT 10
CLR2_MOUSE
ID CLR2_MOUSE STANDARD; PRT; 2920 AA.
AC Q9R0M0; Q9Z2R4; Q99K36;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)
DE (mFml).
```

```
GN CELSR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE-99418630; PubMed-10490098;
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of frizzled.";
RL Cell 98:585-595(1999).
RN [2]
RX SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.
RP PubMed-10790539;
RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;
RT "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is
RT a candidate for the tipy (tip) lethal mutant on chromosome 9.";
RN Mamm. Genome 11:392-394(2000).
RN [3]
RX SEQUENCE OF 2014-2920 FROM N.A.
RP TISSUE-breast tumor;
RA Strausberg R.;
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RX DEVELOPMENTAL STAGE.
RP PubMed-11850187;
RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celsr (Flamingo) genes in the
RT mouse.";
RL Mech. Dev. 112:157-160(2002).
CC -|- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- DEVELOPMENTAL STAGE: Predominantly expressed in the developing
CC CNS, the emerging dorsal root ganglia and cranial ganglia. In the
CC gastrulation, it is expressed within the anterior neural ectoderm.
CC At E10, expression is strong in the ventricular zones (VZ) in all
CC sectors of the brain, and lower in the marginal zones (MZ). It is
CC strong in VZ, lower in MZ, except in telecephalic MZ where it is
CC predominant. The intensity is higher in all VZ, and lower in
CC differentiating fields than in VZ, except in the cerebral
CC hemispheres, and to a lesser extent in the tectum and cerebellum.
CC A weak expression is also observed in the fetal lungs, kidney and
CC epithelia. In the newborn and postnatal stages, expression remains
CC restricted to the VZ as well as in migrating and postmigratory
CC cells throughout the brain.
CC -|- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.
CC -|- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -|- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.
CC -|- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB028499; BAA84070.1; -
CC EMBL; AF031573; AAC68837.1; -
CC EMBL; BC005499; AAH05499.1; -
CC HSPSP; P00740; LEDM
CC MGD; MGI:1858235; Celsr2.
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR002126; Cadherin.
```

DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR000832; GPCR_secretin.
 DR InterPro: IPR001879; hormn_receptor.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR000203; PKD_cys_rich.
 DR Pfam: PF000002; 7tm2; 1.
 DR Pfam: PF000028; cadherin; 9.
 DR Pfam: PF000008; EGF; 5.
 DR Pfam: PF01825; GPS; 1.
 DR Pfam: PF02793; HRM; 1.
 DR Pfam: PF000054; laminin_G; 2.
 DR PRINTS: PR00205; CADHERIN.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PRINTS: PR00249; GPCRSECRETIN.
 DR SMART: SM00112; CA; 9.
 DR SMART: SM00180; EGF_Lam; 1.
 DR SMART: SM00001; EGF_like; 6.
 DR SMART: SM00303; GPS; 1.
 DR SMART: SM00006; HormR; 1.
 DR SMART: SM00282; LamG; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00232; CADHERIN_1; 6.
 DR PROSITE: PS0268; CADHERIN_2; 9.
 DR PROSITE: PS00022; EGF_1; 6.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
 DR PROSITE: PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
 DR PROSITE: PS0227; G_PROTEIN_RECEP_F2_3; 1.
 DR PROSITE: PS0261; G_PROTEIN_RECEP_F2_4; 1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE: PS0221; GPS; 1.
 DR PROSITE: PS00025; LAM_G_DOMAIN; 2.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein;
 DR EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
 DR Developmental protein; Hydroxylation; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 2920
 FT FT
 FT DOMAIN 32 2381
 FT TRANSMEM 2382 2402
 FT DOMAIN 2403 2414
 FT TRANSMEM 2415 2434
 FT DOMAIN 2435 2439
 FT TRANSMEM 2440 2460
 FT DOMAIN 2461 2481
 FT TRANSMEM 2482 2502
 FT DOMAIN 2503 2519
 FT TRANSMEM 2520 2540
 FT DOMAIN 2541 2564
 FT TRANSMEM 2565 2585
 FT DOMAIN 2586 2592
 FT TRANSMEM 2593 2613
 FT DOMAIN 2614 2920
 FT DOMAIN 182 289
 FT DOMAIN 290 399
 FT DOMAIN 400 506
 FT DOMAIN 507 611
 FT DOMAIN 612 713
 FT DOMAIN 714 816
 FT DOMAIN 817 922
 FT DOMAIN 923 1024
 FT DOMAIN 1029 1147
 FT DOMAIN 1229 1287
 FT DOMAIN 1289 1319
 FT DOMAIN 1329 1367
 FT DOMAIN 1368 1572
 FT DOMAIN 1575 1611
 FT DOMAIN 1615 1792
 FT DOMAIN 1788 1830
 FT DOMAIN 1831 1868
 FT DOMAIN 1884 1923

FT DOMAIN 1924 1956
 FT DOMAIN 2317 2369
 FT DOMAIN 2744 2749
 FT DISULFID 1293 1304
 FT DISULFID 1298 1313
 FT DISULFID 1315 1318
 FT DISULFID 1333 1344
 FT DISULFID 1338 1354
 FT DISULFID 1356 1366
 FT DISULFID 1579 1590
 FT DISULFID 1584 1599
 FT DISULFID 1601 1610
 FT DISULFID 1792 1803
 FT DISULFID 1798 1818
 FT DISULFID 1820 1829
 FT DISULFID 1833 1844
 FT DISULFID 1838 1856
 FT DISULFID 1858 1867
 FT DISULFID 1888 1900
 FT DISULFID 1890 1907
 FT DISULFID 1909 1922
 FT DISULFID 1925 1937
 FT DISULFID 1927 1944
 FT DISULFID 1946 1955
 FT MOD_RES 1592 1592
 FT CARBOHYD 486 486
 FT CARBOHYD 558 558
 FT CARBOHYD 702 702
 FT CARBOHYD 1037 1037
 FT CARBOHYD 1077 1077
 FT CARBOHYD 1183 1183
 FT CARBOHYD 1213 1213
 FT CARBOHYD 1502 1502
 FT CARBOHYD 1566 1566
 FT CARBOHYD 1742 1742
 FT CARBOHYD 1828 1828
 FT CARBOHYD 1901 1901
 FT CARBOHYD 2025 2025
 FT CARBOHYD 2044 2044
 FT CARBOHYD 2062 2062
 FT CARBOHYD 2324 2324
 FT CARBOHYD 2346 2346
 FT CONFLICT 2199 2199
 FT CONFLICT 2283 2283
 FT CONFLICT 2535 2535
 FT CONFLICT 2571 2571
 FT CONFLICT 2639 2639

Query Match 84.08; Score 42; DB 1; Length 2920;
 Best Local Similarity 60.08; Pred. No. 25;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
 II::: III
 Db 1820 CVLGYGDNC 1829

RESULT 11
 7B2_XENLA

ID 7B2_XENLA STANDARD; PRT; 161 AA.
 AC P18844;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neuroendocrine protein 7B2 (Secretogranin V) (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_Taxid=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89231705; PubMed=2714283;

FT DISULFID 338 BY SIMILARITY.
 FT DISULFID 349 BY SIMILARITY.
 FT DISULFID 358 BY SIMILARITY.
 FT DISULFID 360 BY SIMILARITY.
 FT DISULFID 375 BY SIMILARITY.
 FT DISULFID 386 BY SIMILARITY.
 FT DISULFID 397 BY SIMILARITY.
 FT DISULFID 399 BY SIMILARITY.
 FT DISULFID 408 BY SIMILARITY.
 FT DISULFID 415 BY SIMILARITY.
 FT DISULFID 422 BY SIMILARITY.
 FT DISULFID 437 BY SIMILARITY.
 FT DISULFID 439 BY SIMILARITY.
 FT DISULFID 448 BY SIMILARITY.
 FT DISULFID 455 BY SIMILARITY.
 FT DISULFID 460 BY SIMILARITY.
 FT DISULFID 475 BY SIMILARITY.
 FT DISULFID 477 BY SIMILARITY.
 FT DISULFID 493 BY SIMILARITY.
 FT DISULFID 498 BY SIMILARITY.
 FT DISULFID 512 BY SIMILARITY.
 FT DISULFID 514 BY SIMILARITY.
 FT DISULFID 523 BY SIMILARITY.
 FT DISULFID 530 BY SIMILARITY.
 FT DISULFID 535 BY SIMILARITY.
 FT DISULFID 552 BY SIMILARITY.
 FT DISULFID 561 BY SIMILARITY.
 FT DISULFID 568 BY SIMILARITY.
 FT DISULFID 578 BY SIMILARITY.
 FT DISULFID 587 BY SIMILARITY.
 FT DISULFID 589 BY SIMILARITY.
 FT DISULFID 605 BY SIMILARITY.
 FT DISULFID 610 BY SIMILARITY.
 FT DISULFID 627 BY SIMILARITY.
 FT DISULFID 636 BY SIMILARITY.
 FT DISULFID 643 BY SIMILARITY.
 FT DISULFID 662 BY SIMILARITY.
 FT DISULFID 664 BY SIMILARITY.
 FT DISULFID 680 BY SIMILARITY.
 FT DISULFID 685 BY SIMILARITY.
 FT DISULFID 702 BY SIMILARITY.
 FT DISULFID 711 BY SIMILARITY.
 FT DISULFID 718 BY SIMILARITY.
 FT DISULFID 723 BY SIMILARITY.
 FT DISULFID 739 BY SIMILARITY.
 FT DISULFID 755 BY SIMILARITY.
 FT DISULFID 760 BY SIMILARITY.
 FT DISULFID 777 BY SIMILARITY.
 FT DISULFID 786 BY SIMILARITY.
 FT DISULFID 793 BY SIMILARITY.
 FT DISULFID 798 BY SIMILARITY.
 FT DISULFID 815 BY SIMILARITY.
 FT DISULFID 824 BY SIMILARITY.
 FT DISULFID 831 BY SIMILARITY.
 FT DISULFID 842 BY SIMILARITY.
 FT DISULFID 855 BY SIMILARITY.
 FT DISULFID 864 BY SIMILARITY.
 FT DISULFID 871 BY SIMILARITY.
 FT DISULFID 876 BY SIMILARITY.
 FT DISULFID 891 BY SIMILARITY.
 FT DISULFID 893 BY SIMILARITY.
 FT DISULFID 902 BY SIMILARITY.
 FT DISULFID 909 BY SIMILARITY.
 FT DISULFID 914 BY SIMILARITY.
 FT DISULFID 931 BY SIMILARITY.
 FT DISULFID 947 BY SIMILARITY.
 FT DISULFID 952 BY SIMILARITY.
 FT DISULFID 957 BY SIMILARITY.
 FT DISULFID 967 BY SIMILARITY.
 FT DISULFID 969 BY SIMILARITY.

Query Match 80.0% Score 40; DB 1; Length 2437;

Best Local Similarity 60.0% Pred. No. 47;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGXSGDXC 10
 |||::|||
 Db 322 CVNGWTGDDC 331

RESULT 15

DL_DROME STANDARD; PRT; 833 AA.
 AC P10041; Q9VDY2; Q99108;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus Delta protein precursor.
 GN DL OR CG3619.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC NCBI_TaxID=7227;
 RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RP TISSUE=Embryo;
 RC VAessin H., Bremer K.A., Knust E., Campos-Ortega J.A.;
 RA "The neurogenic gene Delta of Drosophila melanogaster is expressed in
 RT neurogenic territories and encodes a putative transmembrane protein
 RL with EGF-like repeats";
 RN EMBO J. 6:3431-3440(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R; TISSUE=Embryo;
 RX MEDLINE=89196890; PubMed=3149249;
 RA KOPczynski C.C., Alton A.K., Fechtel K., Kooh P.J., Muskavitch M.A.T.;
 RT "Delta, a Drosophila neurogenic gene, is transcriptionally complex and
 RL encodes a protein related to blood coagulation factors and epidermal
 RN growth factor of vertebrates.";
 RN Genes Dev. 2:1723-1735(1988).
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE OF 422-621 FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=87218537; PubMed=3107986;
 RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D., Vaessin H.,
 RA Campos-Ortega J.A.;
 RT "Egf homologous sequences encoded in the genome of Drosophila
 RL melanogaster, and their relation to neurogenic genes.";
 RN EMBO J. 6:761-766(1987).
 RP [5]
 RN PATTERN OF TRANSCRIPTION, AND CHARACTERIZATION.
 RX MEDLINE=91209246; PubMed=2128477;

RA Haenlin M., Kramatschek B., Campos-Ortega J.A.;
RT "The pattern of transcription of the neurogenic gene Delta of
RT Drosophila melanogaster";
RL Development 110:905-914(1990).
CC -!- FUNCTION: ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM. DL
CC IS REQUIRED FOR THE CORRECT SEPARATION OF NEURAL AND EPIDERMAL
CC CELL LINEAGES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL AREAS WITH NEUROGENIC
CC ABILITIES, FOR EXAMPLE THE NEUROGENIC ECTODERM AND THE PRIMORDIA
CC OF THE SENSE ORGANS. LATER EXPRESSION IS RESTRICTED TO THOSE CELLS
CC THAT HAVE ADOPTED A NEURAL FATE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC EXPRESSION IS HIGHEST EARLY IN EMBRYONIC DEVELOPMENT (STAGE 5) AND
CC REDUCES TO A LOW LEVEL DURING LARVAL STAGES.
CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
CC THE INNER PART OF EMBRIO IS ONE OF THE FIRST STEPS OF CNS
CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
CC NEUROGENIC GENES.
CC -!- MISCELLANEOUS: NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN
CC LEVEL. IT IS CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY
CC COMPETE FOR BINDING WITH THE NOTCH PROTEIN.
CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DSL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X06289; CAA29617.1; -;
DR EMBL; Y00222; CAA68369.1; -;
DR EMBL; AE003725; AAF55657.1; -;
DR EMBL; X05140; CAA28786.1; -;
DR PIR; S00670; S00670.
DR PIR; A26637; A26637.
DR HSR; P00740; LEDM.
DR FlyBase; FBgn0000463; DL.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001774; DSL.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00008; EGF; 9.
DR Pfam; PF01414; DSL; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00051; DSL; 1.
DR SMART; SM00179; EGF_Ca; 4.
DR SMART; SM00001; EGF_like; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00022; EGF_1; 9.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS01187; EGF_Ca; 2.
KW Developmental protein; differentiation; Neurogenesis; Repeat;
KW Transmembrane; EGF-like domain; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 833
FT DOMAIN 19 594
FT TRANSMEM 595 617
FT DOMAIN 618 833
FT CYTOPLASMIC 833 833
FT DSL 226 226
FT DOMAIN 227 258
FT EGF-LIKE 1- 258
FT DOMAIN 256 289
FT EGF-LIKE 2- 289
FT DOMAIN 291 329
FT EGF-LIKE 3- 329
FT DOMAIN 331 372
FT EGF-LIKE 4- 372
FT DOMAIN 374 416
FT EGF-LIKE 5- 416
FT DOMAIN 418 451
FT EGF-LIKE 6- 451

FT	DOMAIN	453	489	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	491	527	EGF-LIKE 8.
FT	DOMAIN	529	565	EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).
FT	DISULFID	231	240	BY SIMILARITY.
FT	DISULFID	235	246	BY SIMILARITY.
FT	DISULFID	248	257	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	266	277	BY SIMILARITY.
FT	DISULFID	279	288	BY SIMILARITY.
FT	DISULFID	295	307	BY SIMILARITY.
FT	DISULFID	301	317	BY SIMILARITY.
FT	DISULFID	319	328	BY SIMILARITY.
FT	DISULFID	335	348	BY SIMILARITY.
FT	DISULFID	342	360	BY SIMILARITY.
FT	DISULFID	362	371	BY SIMILARITY.
FT	DISULFID	378	388	BY SIMILARITY.
FT	DISULFID	383	404	BY SIMILARITY.
FT	DISULFID	406	415	BY SIMILARITY.
FT	DISULFID	422	433	BY SIMILARITY.
FT	DISULFID	427	439	BY SIMILARITY.
FT	DISULFID	441	450	BY SIMILARITY.
FT	DISULFID	457	468	BY SIMILARITY.
FT	DISULFID	462	477	BY SIMILARITY.
FT	DISULFID	479	488	BY SIMILARITY.
FT	DISULFID	495	506	BY SIMILARITY.
FT	DISULFID	500	515	BY SIMILARITY.
FT	DISULFID	517	526	BY SIMILARITY.
FT	DISULFID	533	544	BY SIMILARITY.
FT	DISULFID	538	553	BY SIMILARITY.
FT	DISULFID	555	564	BY SIMILARITY.
FT	CARBOHYD	98	98	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	137	137	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	167	167	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	5	5	K -> N (IN REF. 1).
FT	CONFLICT	67	67	V -> L (IN REF. 1).
FT	CONFLICT	363	363	A -> R (IN REF. 1).
FT	CONFLICT	437	438	GK -> ET (IN REF. 3).
FT	CONFLICT	443	443	A -> S (IN REF. 1 AND 3).
FT	CONFLICT	459	459	G -> A (IN REF. 3).
FT	CONFLICT	490	490	S -> T (IN REF. 3).
FT	CONFLICT	591	591	T -> A (IN REF. 1 AND 3).
FT	CONFLICT	631	631	D -> N (IN REF. 1).
FT	CONFLICT	652	652	G -> A (IN REF. 1).
FT	CONFLICT	662	662	L -> M (IN REF. 1).

Query Match 78.0%; Score 39; DB 1; Length 833;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY	1	CVIGXSGDXC	10
DB	319	CAPGYSGDDC	328

Search completed: July 2, 2003, 07:31:34
Job time : 11.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein : protein search, using sw model
Run on: July 2, 2003, 07:25:01 ; Search time 41.3333 Seconds
(without alignments)
49.850 Million cell updates/sec

Title: US-09-673-785D-13
Perfect score: 50
Sequence: 1 CVIGXGDXC 10

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SP-archaea:*
 - 2: SP-bacteria:*
 - 3: SP-fungi:*
 - 4: SP-human:*
 - 5: SP-invertebrate:*
 - 6: SP-mammal:*
 - 7: SP-mhc:*
 - 8: SP-organalle:*
 - 9: SP-phage:*
 - 10: SP-plant:*
 - 11: SP-rodent:*
 - 12: SP-virus:*
 - 13: SP-vertebrate:*
 - 14: SP-unclassified:*
 - 15: SP-rv:*
 - 16: SP-bacteriap:*
 - 17: SP-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	1200	11	Q8vd07 mus musculu
2	45	90.0	2146	5	Q9VC97 drosophila
3	45	90.0	4006	11	Q35452 mus musculu
4	45	90.0	4114	11	O54796 mus musculu
5	43	86.0	372	5	Q21756 caenorhabdi
6	43	86.0	1162	5	Q8WTF0 halocynthia
7	43	86.0	2531	5	O16004 lytechinus
8	42	84.0	191	5	Q8T521 caenorhabdi
9	42	84.0	191	5	Q8T520 caenorhabdi
10	42	84.0	193	5	Q8T522 caenorhabdi
11	42	84.0	193	5	Q8T523 caenorhabdi
12	42	84.0	615	13	O57409 brachydanio
13	42	84.0	963	5	Q9GPM9 caenorhabdi
14	42	84.0	963	5	Q9GPM9 caenorhabdi
15	42	84.0	1270	5	Q9GPN0 caenorhabdi
16	42	84.0	1531	11	Q9WVB5 mus musculu

17	42	84.0	2920	11	Q9ROM0
18	41	82.0	2352	5	O61240
19	40	80.0	106	4	Q9UN93
20	40	80.0	308	4	Q9UN94
21	40	80.0	366	4	Q9UN95
22	40	80.0	601	10	Q9LYF1
23	40	80.0	794	5	Q8T4P0
24	40	80.0	877	4	Q9H3Q6
25	40	80.0	878	4	Q9GZ22
26	40	80.0	901	4	Q9H3Q7
27	40	80.0	957	4	O14651
28	40	80.0	1217	4	Q9UKW9
29	40	80.0	1530	11	Q9WUG5
30	40	80.0	1531	11	O88279
31	39	78.0	89	6	O28867
32	39	78.0	190	11	Q9JKW3
33	39	78.0	264	5	Q20043
34	39	78.0	265	11	Q9DCP5
35	39	78.0	278	11	Q9QXT5
36	39	78.0	315	2	O33537
37	39	78.0	379	5	Q19922
38	39	78.0	403	5	O18375
39	39	78.0	422	16	P72725
40	39	78.0	426	16	O8Z098
41	39	78.0	470	10	Q9SUL6
42	39	78.0	473	4	O8W29
43	39	78.0	473	4	O12891
44	39	78.0	473	4	O12891
45	39	78.0	559	5	Q9VZ44

ALIGNMENTS

RESULT 1

Q8VD07	Q8VD07	PRELIMINARY;	PRT; 1200 AA.
ID	Q8VD07		
AC	Q8VD07		
DT	01-MAR-2002 (T-EMBLrel. 20, Created)		
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)		
DE	Similar to epidermal growth factor.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=KIDNEY;		
RA	Straussberg R.;		
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC017681; AAH17681.1; -		
DR	InterPro; IPR000152; Asx_Hydroxyl.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR001336; EGF 1.		
DR	InterPro; IPR001881; EGF-Ca.		
DR	InterPro; IPR000033; Ldl_receptor_rep.		
DR	Pfam; PF00058; ldl_recept_b; 7.		
DR	PRINTS; PR00009; EGF_TGF.		
DR	SMART; SM00181; EGF; 9.		
DR	SMART; SM00179; EGF_CA; 8.		
DR	SMART; SM00135; LY; 9.		
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_3.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.		
DR	PROSITE; PS01186; EGF_2; UNKNOWN_6.		
DR	PROSITE; PS01187; EGF_CA; UNKNOWN_3.		
DR	SEQUENCE 1200 AA; 131317 MW; B6A44F7294746476 CRC64;		
SQ			

Query Match 100.0%; Score 50; DB 11; Length 1200;
Best Local Similarity 80.0%; Pred. No. 0.91;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CVIGSGDXC 10
Db 992 CAVGSGDRC 1001

RESULT 2
Q9VC97 PRELIMINARY; PRT; 2146 AA.
AC Q9VC97;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CRB protein.
GN CRB OR CG6383.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.M., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003747; AAF56276.1; -.
DR HSSP; P00740; IEDM.
DR FlyBase; FBgn0000368; crb.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR002049; Laminin_EGF.
DR DR Pfam; PF0001791; Laminin_G.
DR Pfam; PF00008; EGF; 26.
DR Pfam; PF00054; laminin_G; 3.

```

```

DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00179; EGF_CA; 11.
DR SMART; SM00001; EGF_Like; 16.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 16.
DR PROSITE; PS00022; EGF_1; 25.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 13.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2146 AA; 233570 MW; 8E23B9E32B761115 CRC64;

Query Match 90.0%; Score 45; DB 5; Length 2146;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
Db 569 CAVGSGDRC 578

RESULT 3
Q35452 PRELIMINARY; PRT; 4006 AA.
AC Q35452;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Tenascin X.
GN TNX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA Loretz C., Schmidt S., Traicoff R., Zackrone K., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class III
region."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030001; AAB82015.1; -.
DR HSSP; P02671; IFZD.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00008; EGF; 15.
DR Pfam; PF00147; fibrinogen_C; 1.
DR Pfam; PF00041; fn3; 30.
DR SMART; SM00181; EGF; 8.
DR SMART; SM00001; EGF_Like; 2.
DR SMART; SM00186; FBG; 1.
DR SMART; SM00060; FN3; 24.
DR PROSITE; PS00022; EGF_1; UNKNOWN_19.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS00514; FIBRIN_AG_DOMAIN; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 4006 AA; 435471 MW; 553PBE873498A4FC CRC64;

Query Match 90.0%; Score 45; DB 11; Length 4006;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGSGDXC 10
Db 532 CAVGSGDRC 541

RESULT 4
O54796 PRELIMINARY; PRT; 4114 AA.
ID O54796
AC O54796;
DT 01-JUN-1998 (TREMBLrel. 06, Created)

```



```

RESULT 7
OL6004 ID O16004 PRELIMINARY; PRT; 2531 AA.
AC O16004;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Notch homolog.
OS Lytechinus variegatus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinozoa; Euechinozoa; Echinacea; Temnopleuroidea; Toxopneustidae;
OX NCBI_TaxID=7654;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97454256; PubMed=9310331;
RA Sherwood D.R., McClay D.R.;
RT "Identification and localization of a sea urchin Notch homologue:
RT insights into vegetal plate regionalization and Notch receptor
RT regulation.";
RL Development 124:3363-3374(1997).
DR EMBL; AF000634; AAB82088.1; -
DR HSP; P01132; IEGF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; ASX_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PRO1415; ANKYRIN.
DR PRINTS; PRO0010; EGF_BLOOD.
DR PRINTS; PRO1452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00001; EGF_like; 11.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 10.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 21.
DR PROSITE; PS00022; EGF_1; UNKNOWN_33.
DR PROSITE; PS01186; EGF_2; 25.
DR PROSITE; PS01187; EGF_CA; 20.
KW ANK repeat; Calcium-binding; EGF-like domain; Glycoprotein;
KW Hydroxylation; Repeat.
SQ SEQUENCE 2531 AA; 273982 MW; 5BF42BEC627CA303 CRC64;

Query Match 86.0%; Score 43; DB 5; Length 2531;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
DB 615 CPVGTSGDNC 624
I : : : : :

RESULT 8
Q8T521 ID Q8T521 PRELIMINARY; PRT; 191 AA.
AC Q8T521;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB800;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491462; AAM09702.1; -
FT NON_TER 1
FT NON_TER 191
SQ SEQUENCE 191 AA; 20977 MW; 2EEB21B5FFA46470 CRC64;

Query Match 84.0%; Score 42; DB 5; Length 191;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
DB 53 CPLGYSGDYC 62
I : : : : :

RESULT 9
Q8T520 ID Q8T520 PRELIMINARY; PRT; 191 AA.
AC Q8T520;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VT847;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).
DR EMBL; AF491463; AAM09703.1; -
FT NON_TER 1
FT NON_TER 191
SQ SEQUENCE 191 AA; 20943 MW; 2EEB21B5F47FB470 CRC64;

Query Match 84.0%; Score 42; DB 5; Length 191;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDNC 10
DB 53 CPLGYSGDYC 62
I : : : : :

RESULT 10
Q8T522 ID Q8T522 PRELIMINARY; PRT; 193 AA.
AC Q8T522;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Glp-1 (Fragment).
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF16;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
RT genus Caenorhabditis.";
RL Genetics 0:0-0(2002).

```

```

DR EMBL; AF491460; AAM09700.1; -.
FT NON_TER 1
FT NON_TER 193
SQ SEQUENCE 193 AA; 21178 MW; 42135BB8E8BE02C5 CRC64;

Query Match      84.0%; Score 42; DB 5; Length 193;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   | : : : : :
Db 55 CPLGYSGDYC 64

RESULT 11
Q8STG0 PRELIMINARY; PRT; 193 AA.
AC Q8STG0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Glp-1 (Fragment)
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK104, AND HK105;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
   genus Caenorhabditis."
RL Genetics 0:0-0(2002).
DR EMBL; AF491459; AAM09699.1; -.
DR EMBL; AF491461; AAM09701.1; -.
FT NON_TER 1
FT NON_TER 193
SQ SEQUENCE 193 AA; 21180 MW; 413C8AB647B5C540 CRC64;

Query Match      84.0%; Score 42; DB 5; Length 193;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   | : : : : :
Db 56 CPLGYSGDYC 65

RESULT 12
Q8T5Z3 PRELIMINARY; PRT; 194 AA.
AC Q8T5Z3;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Glp-1 (Fragment)
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB826;
RA Graustein A., Gaspar J.M., Walters J.R., Palopoli M.F.;
RT "Levels of DNA polymorphism vary with mating system in the nematode
   genus Caenorhabditis."
RL Genetics 0:0-0(2002).
DR EMBL; AF491458; AAM09698.1; -.
FT NON_TER 1
FT NON_TER 194
SQ SEQUENCE 194 AA; 21275 MW; 07313C8AB9C65C5 CRC64;

Query Match      84.0%; Score 42; DB 5; Length 194;

```

```

Best Local Similarity 60.0%; Pred. No. 4.8;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   | : : : : :
Db 56 CPLGYSGDYC 65

RESULT 13
O57409 PRELIMINARY; PRT; 615 AA.
AC O57409;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Deltab.
GN DLB OR DELTAB.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165391; PubMed=9425132;
RA Haddon C., Smithers L., Schneider-Maunoury S., Coche T., Henrique D.,
   Lewis J.;
RT "Multiple delta genes and lateral inhibition in zebrafish primary
   neurogenesis."
RL Development 125:359-370(1998).
DR EMBL; AF006488; AAC41241.1; -.
DR HSP; P00740; 1EDM.
DR 2FIN; ZDB-GENE-980526-114; dlb.
DR InterPro; IPR00152; Asx_hydroxyl.
DR InterPro; IPR001774; DSL.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR Pfam; PF01414; DSL; 1.
DR Pfam; PF00088; EGF; 7.
DR PRINTS; PR00010; EGFLOOD.
DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00001; EGF_like; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 615 AA; 67592 MW; CA18004428F5603C CRC64;

Query Match      84.0%; Score 42; DB 13; Length 615;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
   | : : : : :
Db 194 CLVGWQGDYC 203

RESULT 14
O9GPM9 PRELIMINARY; PRT; 963 AA.
AC O9GPM9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Notch-like transmembrane receptor (Fragment).
GN GLP-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

```
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AF16;
RX MEDLINE=21100341; PubMed=11156985;
RA Rudel D., Kimble J.;
RT "Conservation of glp-1 regulation and function in nematodes.";
RL Genetics 157:639-654(2001).
DR EMBL; AF315555; AAG49317.1; -.
DR HSSP; P01132; IEGF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00181; EGF; 5.
DR SMART; SM00001; EGF_like; 4.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_4.
DR PROSITE; PS01186; EGF_2; 3.
DR ANK repeat; Receptor; Repeat; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 963 AA; 106394 MW; 52BF80010E12FB57 CRC64;

Query Match 84.0%; Score 42; DB 5; Length 963;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 123 CPLGYSGDYC 132
|.:|:|:|

RESULT 15
Q9GPN0
ID Q9GPN0 PRELIMINARY; PRT; 1270 AA.
AC Q9GPN0;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Notch-like transmembrane receptor.
GN GLP-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AF16;
RX MEDLINE=21100341; PubMed=11156985;
RA Rudel D., Kimble J.;
RT "Conservation of glp-1 regulation and function in nematodes.";
RL Genetics 157:639-654(2001).
DR EMBL; AF315554; AAG49316.1; -.
DR HSSP; P01132; IEGF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001891; EGF_Ca.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
```

```
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00181; EGF; 12.
DR SMART; SM00179; EGF_CA; 10.
DR SMART; SM00001; EGF_like; 8.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 1.
DR ANK repeat; Receptor; Repeat; Transmembrane.
SQ SEQUENCE 1270 AA; 138964 MW; A7662EB575A4B61B CRC64;

Query Match 84.0%; Score 42; DB 5; Length 1270;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGXSGDXC 10
Db 430 CPLGYSGDYC 439
|.:|:|:|

Search completed: July 2, 2003, 07:28:39
Job time : 42.3333 secs
```